PERIOD-LUMINOSITY RELATIONS OF CEPHEID AND MIRA VARIABLES AND THEIR APPLICATION TO THE EXTRAGALACTIC DISTANCE SCALE

A Dissertation

by

WENLONG YUAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Lucas M. Macri
Committee Members,	Jianhua Z. Huang
	Jennifer L. Marshall
	Nicholas B. Suntzeff
Head of Department,	Peter McIntyre

August 2017

Major Subject: Physics

ABSTRACT

In this dissertation, I present work towards accurate and precise distance determinations using both Cepheids and Miras. The work includes a Cepheid search in the galaxy M101, a study of near-infrared light curves and phase corrections for Galactic Cepheids, a Mira search in the galaxy M33, a study of near-infrared properties of Miras in the Large Magellanic Cloud, and an investigation into the suitability of the Large Synoptic Survey Telescope for detecting extragalactic Miras. Using time-series observations of two fields in M101, we identified hundreds of Cepheids and derived their mean magnitudes We obtained optical Cepheid Period-Luminosity Relations and a reddening-corrected distance to this galaxy. Combining the ground-based time-series observations for 34 Galactic Cepheids with literature data, we determined the contemporary phases of a sample of Galactic Cepheids. We used the phases and light curves to correct the single-epoch space-based observations to their mean values. Collaborating with statisticians, we carried out a Mira search in the Local Group galaxy M33 by coupling a novel semi-parametric Gaussian process model and machine learning techniques. We discovered 1847 Mira candidates using I-band measurements and obtained preliminary Period-Luminosity Relations at multiple wavelengths. We studied the near-infrared properties of Large Magellanic Cloud Mira candidates using multi-epoch JHK_s observations. We found that the color excesses of Oxygen- and Carbon-rich Miras are different and compared them with the interstellar extinction law. We obtained the near-infrared Mira Period-Luminosity Relations for the Oxygen-rich subtype. We investigated the feasibility of discovering Miras with the Large Synoptic Survey Telescope. We found that our method will discover a considerable number of Oxygen-rich Miras in dozens of systems within 15 Mpc.

ACKNOWLEDGMENTS

I would like to thank my PhD Advisor, Lucas Macri, for his mentorship and support during these past five years. He always gives instant yet comprehensive answers to my questions, and insightful suggestions for my research. I thank my collaborators in the Department of Statistics at Texas A&M University, Shiyuan He, Jianhua Huang, and James Long and all the members in my PhD committee. I would also like to thank my parents, sister, and wife for their constant love and support.

This research has made use of the following resources:

- the Mikulski Archive for Space Telescopes (MAST) at STScI, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.
- data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.
- data products from the Optical Gravitational Lensing Experiment, conducted by the Astronomical Institute of the University of Warsaw at Las Campanas Observatory, operated by the Carnegie Institution for Science.
- the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.
- the variable star observations from the AAVSO International Database contributed by observers worldwide.

- the VizieR catalogue access tool, CDS, Strasbourg, France.
- the Texas A&M University Brazos HPC cluster.
- NASA's Astrophysics Data System.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Lucas Macri (advisor), Jennifer Marshall and Nicholas Suntzeff of the Department of Physics and Astronomy and Professor Jianhua Huang of the Department of Statistics.

The analyses depicted in Section 2 were conducted in part by Samantha Hoffmann of the Department of Physics and Astronomy and were published in 2016 in an article listed in the Astrophysical Journal. The analyses depicted in Section 4 were conducted in part by Lucas Macri of the Department of Physics and Astronomy and Shiyuan He, James Long, Jianhua Huang of the Department of Statistics and were published in 2017 in an article listed in the Astronomical Journal.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the National Science Foundation through grant AST-1211603, by the National Aeronautics and Space Administration and the Space Telescope Science Institute through grants HST-GO-12880, 13334, 13335 & 13678, and by the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University.

NOMENCLATURE

2MASS	Two Micron All Sky Survey
ACS	Advanced Camera for Surveys
AGB	Asymptotic Giant Branch
AUC	Area Under the Curve
C-rich	Carbon-rich
CTE	Charge Transfer Efficiency
CTIO	Cerro Tololo Inter-American Observatory
FLWO	Fred L. Whipple Observatory
GR	General Relativity
H_0	Hubble constant
HR diagram	Hertzsprung-Russell diagram
HST	*Hubble Space Telescope*
LMC	Large Magellanic Cloud
LMCNISS	LMC Near-Infrared Synoptic Survey
LSST	Large Synoptic Survey Telescope
MAST	Mikulski Archive for Space Telescopes
NIR	Near-Infrared
OGLE	Optical Gravitational Lensing Experiment
oPLCR	observed Period-Luminosity-Color Relation
O-rich	Oxygen-rich
PLR	Period-Luminosity Relation

PSF	Point Spread Function
RF	Random Forest
ROC	Receiver Operating Characteristic
RSS	Residual Sum of Squares
SMC	Small Magellanic Cloud
SNe Ia	Type Ia Supernovae (plural)
SN Ia	Type Ia Supernova
SNR	Signal-to-Noise Ratio
SRV	Semi-Regular Variable
WFC	Wide Field Camera
WFC3	Wide Field Camera 3
ΛCDM	Lambda Cold Dark Matter

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Since the discovery of the expanding Universe by Hubble (1929), astronomers have endeavored to improve the measurement of its expansion rate, known as the Hubble constant (H_0), in terms of accuracy and precision. H_0 is defined as the apparent recession velocity of an object located at unit distance in the present-day Universe. The state-of-the-art method to measure H_0 is through a cosmic distance ladder, calibrating each rung using overlapping distance indicators. Geometric methods, including trigonometric parallaxes (Casertano et al., 2016), eclipsing binary star systems (Pietrzyński et al., 2013), and water masers around NGC 4258 (Humphreys et al., 2013), provide absolute distances to anchors within $\sim$ 7 Mpc. The Cepheid Period-Luminosity Relation (PLR) is calibrated with these anchors and provides reliable distances to star-forming galaxies up to 50 Mpc away. Within the volume that Cepheid PLRs can reach, the absolute luminosity type Ia supernovae (SNe Ia) is calibrated and applied to a distant sample of SNe Ia, from which H_0 can be measured.

While Cepheids have been extensively used to measure H_0 in the past decades (Freedman et al., 2001; Riess et al., 2009, 2011, 2016), Miras have not been widely used in this area. This is partially due to the fact that Miras are long-period variables with somewhat erratic light curves, making them difficult to identify within short observational baselines or a limited number of epochs. However, with upcoming long-baseline all-sky surveys and near-infrared (NIR) facilities, Miras will be promising alternatives for Cepheids. Miras are $\sim$ 2 mag brighter at NIR wavelengths and ubiquitous in both early- and late-type galaxies. As shown in this dissertation, their PLRs exhibit similar dispersions as Cepheids at NIR wavelengths.

1.1 Cepheids

Classical Cepheid variables, often referred to as Cepheids, are massive luminous yellow supergiant stars in the final stages of stellar evolution. This class of stars was established in the late 18th century when the English astronomers Edward Pigott and John Goodricke discovered the variability of η Aquilae and δ Cephei. Using modern photometric measurements, Cepheids are easily recognized by their sawtooth-shape light curves. For most Cepheids, the light curves are exactly the same from cycle to cycle, which means that densely-sampled smooth curves can be obtained by folding the observed light curve into phase space, as shown in Figure 1.1. The initial masses of Cepheids range from 4 $M_{\odot}$ to 20 $M_{\odot}$ (Turner, 1996), and the sizes are usually between 30 $R_{\odot}$ and 300 $R_{\odot}$ (Gieren et al., 1999). The absolute magnitude of Cepheids is correlated with period, as famously discovered by Leavitt & Pickering (1912), and ranges from -1.5 mag to -7 mag in the V-band (Soszyński et al., 2008).

Cepheids are pulsating stars located in a narrow "instability strip" in the Hertzsprung-Russell (HR) diagram. The effective (photospheric) temperature of a star begins to drop when the hydrogen fuel is exhausted in its core, and its position in the HR diagram moves towards the low-temperature end. The exact evolutionary track depends on the mass and metallicity of the star. Figure 1.2 shows the post-main sequence tracks of stars with solar metallicity and various initial masses based on the model from Schaller et al. (1992). When the stars evolve across the "instability strip" (dashed lines in Figure 1.2), they become unstable as the stellar atmosphere temperature becomes close to the ionization temperature of He I $\rightarrow$ He II. During this transition, the opacity changes dramatically. The release of stellar energy is controlled by the atmosphere and the star begins to pulsate.

Cepheids may pulsate in several modes. Fundamental and first-overtone are the most common modes, while second-overtone and mixed modes are relatively rare (Soszyński

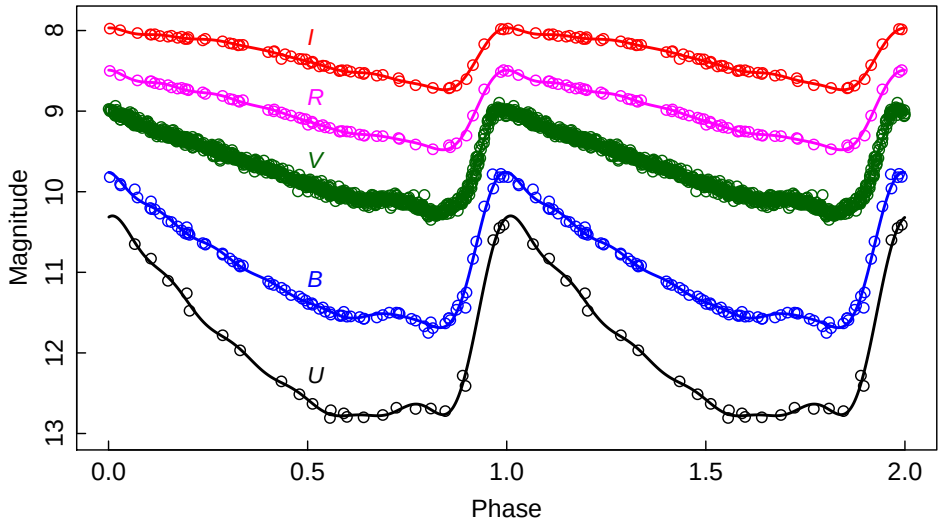

Figure 1.1: Phase-folded light curves for a classical Cepheid, VZ Pup, in $UBVRI$ bands. The measurements were compiled from the McMaster Cepheid Photometry and Radial Velocity Data Archive maintained by Doug Welch and the All Sky Automated Survey (Pojmanski, 1997). The solid lines represent best-fit curves with seventh-order Fourier series.

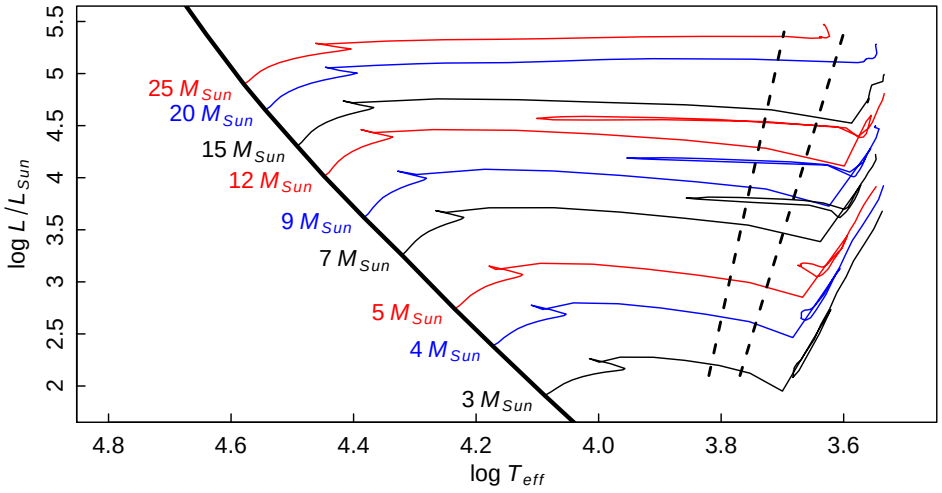

Figure 1.2: Evolutionary tracks of stars with various initial masses in the HR diagram based on Schaller et al. (1992, thin solid curves) and the edges of the fundamental-mode instability strip that are estimated based on the results of Alibert et al. (1999, dashed lines). The thick solid line indicates the location of the main sequence. The metallicity for the stellar models, as welll as for the edges of the instability strip, is set close to solar metallicity ($Z = 0.02$). Part of the material in this plot was reprinted from Schaller et al. (1992).

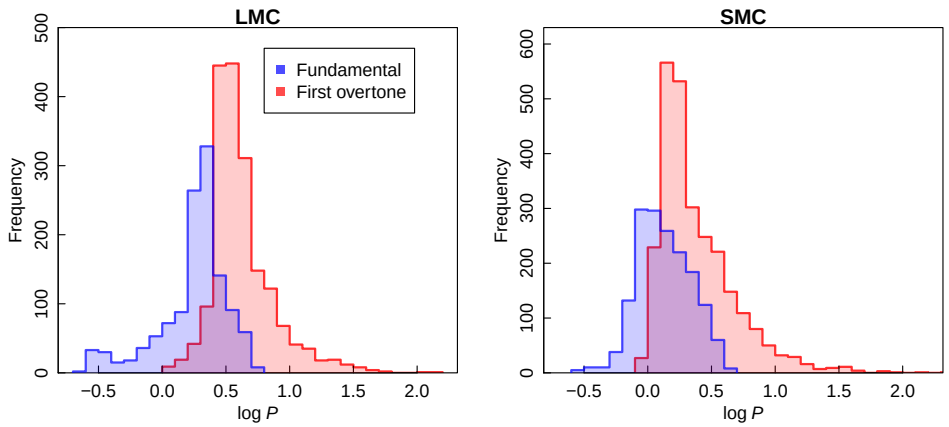

Figure 1.3: Distribution of periods for fundamental-mode (red) and first-overtone (blue) Cepheids in the LMC (left) and the SMC (right).

et al., 2008). Figure 1.3 shows histograms of periods for fundamental-mode Cepheids and first-overtone Cepheids in the Large and Small Magellanic Clouds (LMC, SMC), two fairly complete samples that were collected by the Optical Gravitational Lensing Experiment (OGLE, Udalski et al., 2008). While the periods of fundamental-mode Cepheids can reach beyond 100 days, those of first-overtone Cepheids are limited to $\lesssim 7.5$ days (Baranowski et al., 2009) due to the effects of convection (Smolec & Moskalik, 2008). We used fundamental-mode Cepheids to calibrate the Extragalactic Distance Scale as they reach significantly longer periods and brighter absolute magnitudes.

The Cepheid PLR was discovered by Henrietta S. Leavitt in 1908 when she investigated the variable stars in the LMC and the SMC (Leavitt & Pickering, 1912). It is the linear correlation between the log of periods and the magnitudes of Cepheids. Figure 1.4 shows the Cepheid PLRs of the LMC sample at various optical (Soszyński et al., 2008) and NIR (Macri et al., 2015; Persson et al., 2004) wavelengths. This relation is the natural result of Stefan's law in logarithmic form (Madore & Freedman, 1991):

$$M = -2.5 \log(4\pi R^2 \sigma T_e^4) + C = -5 \log R - 10 \log T_e + C \,, \tag{1.1}$$

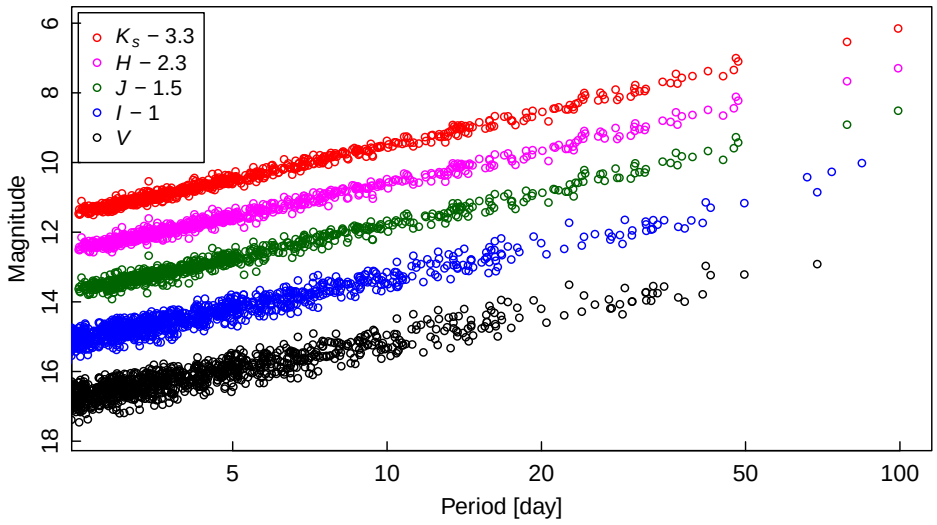

Figure 1.4: Cepheid PLRs in the LMC in the $VIJHK_s$ bands. The VI measurements are obtained from OGLE (Soszyński et al., 2008) while the JHK_s measurements are reprinted from Macri et al. (2015) and Persson et al. (2004).

where M is the bolometric magnitude, R is the radius, and T_e is the effective temperature. Under the assumption of an ideal gas law, the radial pulsation period for adiabatic process (Collins, 1989)

$$P = (\frac{4\pi}{3G\bar{\rho}})^{1/2} \sim R^{3/2} \tag{1.2}$$

predicts the PLR in the form of

$$M = a \log P + f(T) \tag{1.3}$$

where $f(T)$ is the temperature dependence and the coefficient $a = -10/3$. The Cepheid PLR serves as powerful tool to measure distances.

1.2 Miras

Miras are long-period pulsating asymptotic giant branch (AGB) stars that exhibit large cyclical variations in flux at optical wavelengths. The "canonical" empirical classification requires $\Delta V > 2.5$ mag within a pulsation cycle and spectroscopic confirmation (Kholopov et al., 1985). Recent surveys for these variables (such as Soszyński et al., 2009) have adopted $\Delta I > 0.8$ mag as the requirement for classification, since spectroscopic follow-up of very large samples is not currently feasible. Longer-term variations in the mean flux level of each cycle are typical (Mattei, 1997; Whitelock et al., 1997), and visual light curves exhibit a wide range of shapes; Ludendorff (1928) classified Miras into various classes and subclasses based on this attribute. The origin of these variations is not clear so far. The initial masses of Miras range from $\sim 1M_{\odot}$ to $\sim 4M_{\odot}$ (Feast, 2009), but the mass loss during the AGB phase, which forms the planetary nebulae, can be considerable (Wood, 1990). The sizes of Miras are typically between 250 $R_{\odot}$ and 800 $R_{\odot}$ (van Belle et al., 2002), and they are more luminous at NIR wavelengths than in the optical, with K_s-band absolute magnitudes between -6 mag to -11 mag depending on their periods. The origin of their pulsation mechanism is not currently fully understood, but opacity changes due to the transition H $\rightarrow$ H I is a likely cause.

The prototype of the Mira class is o Ceti, whose variability was first recorded by the German astronomer David Fabricius in the late 16th century. o Ceti is a long-period (332 days) variable with a visual magnitude changing cyclically from 10 mag to 2 mag and located at a distance of only ~ 90 pc (van Leeuwen, 2007). Its visual brightness has been recorded for more than a century (Templeton & Karovska, 2009). Figure 1.5 shows the light curve of o Ceti from January, 2010 to April, 2017 in the visual band using the observations compiled by the American Association of Variable Star Observers.

Since the progenitors of Miras are relatively low-mass stars, they are ubiquitous and

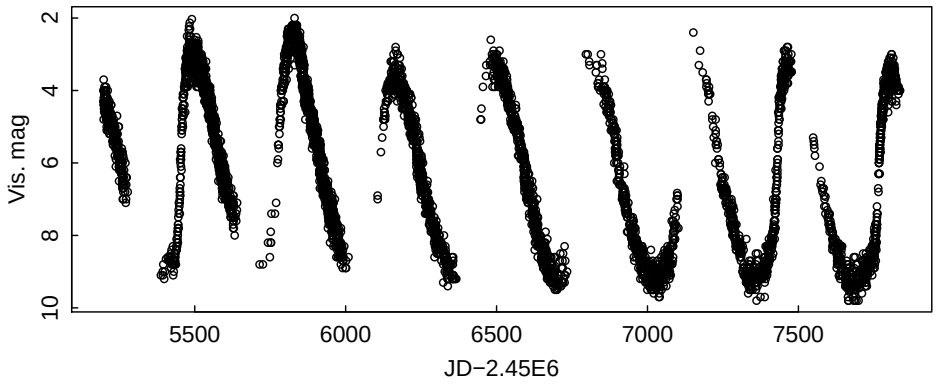

Figure 1.5: The visual light curve of o Ceti from January, 2010 to April, 2017. The data were retrieved from the American Association of Variable Star Observers.

present in all types of galaxies. Thousands of Mira candidates have been discovered in the Milky Way and the Magellanic Clouds based on photometry from the OGLE (Soszyński et al., 2009) and MACHO (Alcock et al., 1993) projects. Recently, large samples of extragalactic Mira candidates have been identified in M33 (Yuan et al., 2017) and NGC 4258 (Huang et al., 2017, in preparation). In just a few years, the Large Synoptic Survey Telescope (LSST) will begin to obtain frequent images of dozens of nearby galaxies that will have the necessary depth to enable the discovery of Miras and the determination of distances to these systems (see discussion in Section 6).

Miras are divided into two subtypes: Oxygen- and Carbon-rich (O-rich, C-rich), based on which of these elements is more abundant in their atmospheres. Spectroscopically, the O-rich subtype features emissions at 10 μm, 11 μm, and 13 μm possibly due to amorphous silicate and alumina (Miyata et al., 2000), while the C-rich subtype is characterized by broad emissions which are likely related to amorphous carbon, SiC, polycyclic aromatic hydrocarbons, C_{60}, and C_{70} (Mutschke et al., 1999; Cami et al., 2010). The stable CO molecules lock equal amounts of carbon and oxygen and only the element of greater abundance is left for further chemical reactions, which leads to the chemical dichotomy of

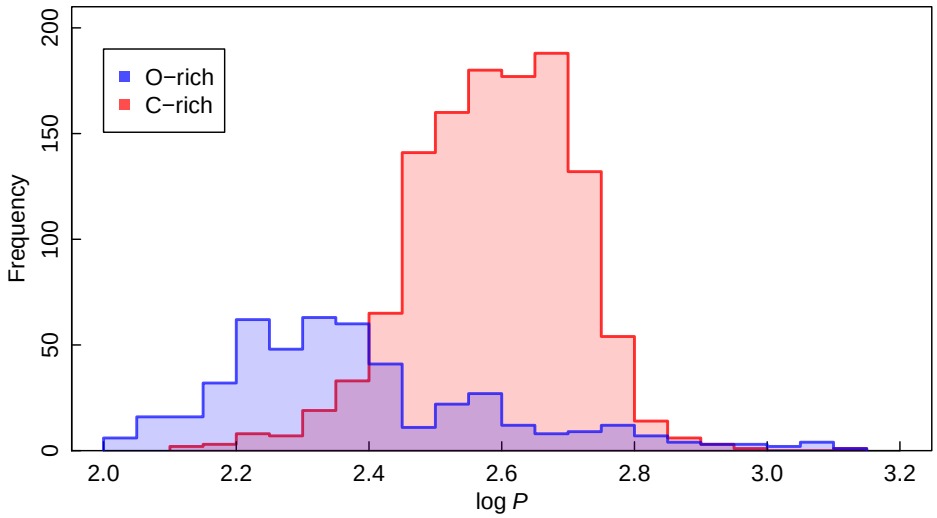

Figure 1.6: Distribution of periods for C- (red) and O-rich Miras (blue) in the LMC.

O- and C-rich subtypes. A number of intermediate-type Miras were also found to exhibit both O- and C-rich features (e.g. Guzman-Ramirez et al., 2015), which suggests an evolutionary path from O- to C-rich subtype as more carbon is dredged up from the interior of the star. Figure 1.6 shows the distribution of periods for these two Mira subtypes in the LMC using the OGLE-III catalog (Soszyński et al., 2009).

The Mira PLR was initially studied by Gerasimovic (1928), who related the periods and mean visual magnitudes at maximum for ten nearby long-period variables that had trigonometric parallaxes (van Maanen, 1928). Glass & Lloyd Evans (1981) found the first evidence of a PLR at NIR wavelengths for Miras, based on a small sample of O-rich variables in the LMC. Using the MACHO database, Wood et al. (1999) were the first to identify multiple PLRs for AGB stars and to confirm the nature of Miras as radial fundamental-mode pulsators. Also using MACHO periods, Glass & Lloyd Evans (2003) determined that the Mira K-band PLR exhibits a relatively small scatter of $\sigma \sim 0.13$ mag, while Whitelock et al. (2008) found similar dispersions for K-band PLRs separated into O-

and C-rich subtypes (σ = 0.14 and 0.15 mag, respectively). These values are comparable to the intrinsic dispersion of the Cepheid PLR in the same bandpass (σ = 0.09 mag; Macri et al., 2015). Soszyński et al. (2009, 2011, 2013) characterized the NIR Mira PLRs in the LMC, the SMC, and the Galactic Bulge using OGLE and Two Micron All Sky Survey (2MASS) photometry. Given their large numbers and higher luminosities in NIR, Miras could be competitive distance indicators in the future.

1.3 Observational Cosmology

Hubble (1929) published an empirical relation between distance D and apparent radial velocity v among galaxies in the local Universe, which can be represented by a linear dependence

$$v = H_0 \cdot D \tag{1.4}$$

where the coefficient H_0 is a positive number and was later named as the Hubble constant. The relation is now called Hubble's Law. Thanks to the development of cosmological theories at that time, such as General Relativity (GR; Einstein, 1916), the Friedmann equations (Friedmann, 1922), and the idea of the Big Bang (Lemaître, 1931), Hubble's Law was interpreted as space itself expanding rather than that remote galaxies receding through space.

In the framework of GR, Alexander Friedmann assumed a homogeneous and isotropic universe with the metric

$$\mathrm{d}s^2 = a(t)^2 \mathrm{d}r^2 - c^2 \mathrm{d}t^2 \tag{1.5}$$

where $a(t)$ is a scale factor evolving with time, and solved the Einstein field equations. The expansion rate of the Universe over cosmological time can be expressed as

$$H = \frac{\dot{a}}{a} \tag{1.6}$$

and its current value is H_0. Hubble's Law suggests an evolving Universe where the cosmological constant Λ, introduced by Einstein in his field equations, is not a "fine-tuned" number to keep the Universe static.

In 1998, two teams independently discovered that the expansion of the Universe is accelerating by measuring the magnitudes and redshifts of distant SNe Ia (Riess et al., 1998; Perlmutter et al., 1999). This new observational fact led to the hypothesis of dark energy, which has a negative pressure and causes the accelerating expansion of the Universe. Thus, the standard cosmological model is now called the Lambda cold dark matter (ΛCDM) model.

Currently, the ΛCDM model is by no means complete. The *Planck* mission inferred a value of $H_0 = 67.8 \pm 0.9$ km s^{-1} Mpc^{-1} based on temperature and lensing measurements under the assumption of ΛCDM (Planck Collaboration et al., 2016), while the direct measurement of H_0 using a classical distance ladder gives 73.2 ± 1.7 km s^{-1} Mpc^{-1}, a value 3.4σ higher (Riess et al., 2016). The calibration of the Extragalactic Distance Scale is still of critical importance in probing and developing cosmological models. In this dissertation, we present several contributions to this effort using the PLRs of Cepheids and Miras.

2. A SEARCH FOR CEPHEIDS IN M101*

As part of an effort to measure H_0 with a total uncertainty of 2.4% (Riess et al., 2016), we obtained *Hubble Space Telescope* (*HST*) observations of the galaxy M101, which hosted the type Ia supernova 2011fe. We combined the new images with archival data and identified over 800 Cepheid candidates with periods ranging from 3 to 80 days. We present the analysis and results of this work, including optical Period-Luminosity Relations (PLRs) and a Cepheid distance to this galaxy.

2.1 Observations, Data Reduction, and Photometry

2.1.1 Observations

Two fields within M101 were observed with *HST* in late 2006 and early 2007 (GO-10918, PI Freedman, hereafter 06/07 observations) in the $F555W$ and $F814W$ bands using the Advanced Camera for Surveys/Wide Field Camera (ACS/WFC). In 2013, we obtained additional $F555W$ and $F814W$ observations as well as $F160W$ observations of the same fields (GO-12880, hereafter 2013 observations) using the Wide Field Camera 3 (WFC3). The 2013 observations covered about 65% of the area of the 06/07 observations. The positions of the observed fields are shown in Figure 2.1. Black boxes indicate the 06/07 observations; blue boxes are the 2013 optical observations; red boxes are the 2013 $F160W$ observations. In the rest of this section we refer the left black field as ACS-F1, the right black field as ACS-F2, and the left blue/red field as WFC3-F1, the right blue/red field as WFC3-F2. Table 2.1 summarizes all the observations used in this work.

The 06/07 observations consist of 12 epochs of $F555W$ and $F814W$ observations for both ACS-F1 and ACS-F2. Images were taken in the form of cosmic-ray splits and there

*Part of the material in this section has published in "Optical Identification of Cepheids in 19 Host Galaxies of Type Ia Supernovae and NGC 4258 with the Hubble Space Telescope" by Hoffmann et al. (2016), *The Astrophysical Journal*, 830:10 (15pp). Copyright 2016 by the American Astronomical Society.

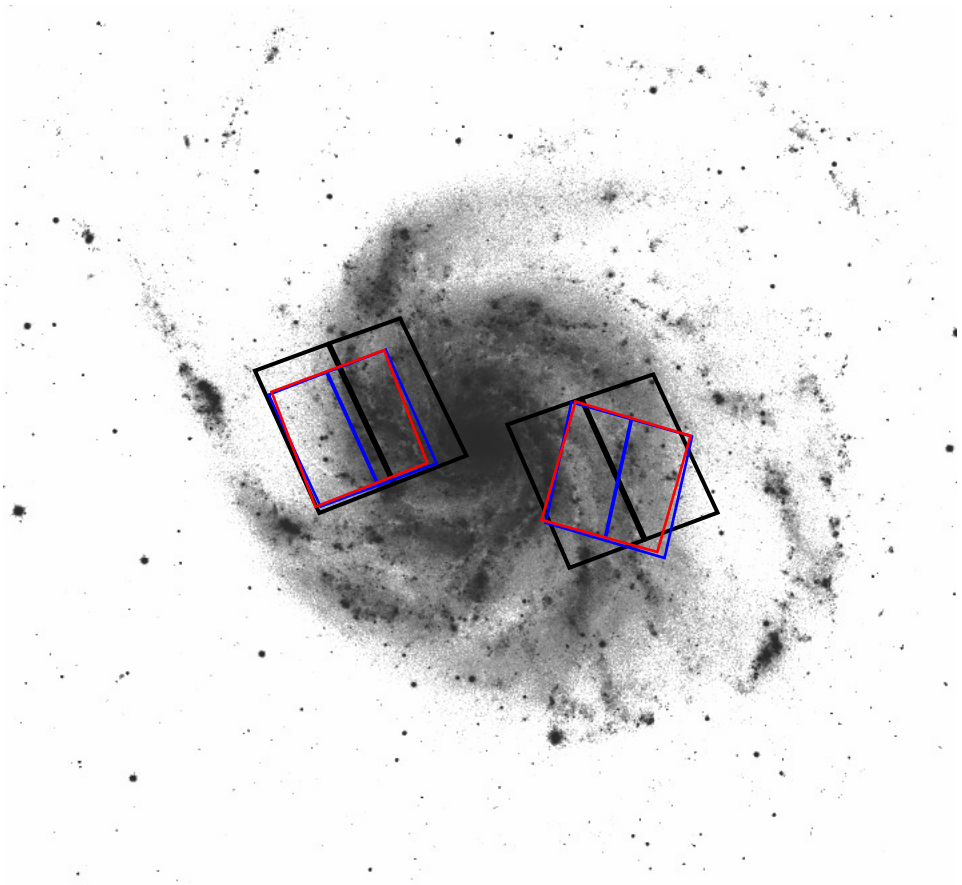

Figure 2.1: The positions of all the observations superposed on Sloan Digital Sky Survey mosaic image of M101. The black boxes, blue boxes, and red boxes are corresponding to 06/07 $F555W$-band and $F814W$-band observations, 2013 $F555W$-band and $F814W$-band observations, and 2013 $F160W$-band observations, respectively. North is up and east is to the left.

are 96 images in total. For each visit, four images were taken in the same orbit, in a sequence of 300s in $F814W$, 700s in $F555W$, 424s in $F814W$, and 630s in $F555W$. The sampling of the first 11 epochs follows a non-redundant distribution, while epoch 12 is off the pattern to keep the observation baseline within 30 days.

The 2013 WFC3 observations consist of 2 epochs of $F555W$, $F814W$, and $F160W$ imaging. They were acquired as single images rather than in cosmic-ray splits. Due to the smaller field of view of WFC3/IR, four $F160W$ images were taken at each epoch using

spatial offsets to cover the WFC3/UVIS field of view. For each visit, four 303s $F160W$, one 472s $F555W$, and one 472s $F814W$ images were taken in the same orbit. The time interval between the two epochs is ~ 6.6 days.

2.1.2 Data Reduction

We retrieved the data from the Mikulski Archive for Space Telescopes (MAST). Due to the differences of the archival and new datasets, we address their reduction procedures separately. The analysis of the $F160W$ observations was excluded from this work.

The images taken with ACS (06/07 observations) were pre-reduced by the On-The-Fly-Reprocessing `calacs` procedure, which performed bias subtraction, dark subtraction, flat fielding, and pixel-based Charge Transfer Efficiency (CTE) correction. We used the `PyRAF DrizzlePac` package (Gonzaga & et al., 2012), `DAOPHOT`, `ALLSTAR`, `DAOMATCH`, and `DAOMASTER` (Stetson, 1987, 1993) to remove the geometric distortions, reject cosmic rays and hot pixels, subtract the sky, and register the frames of each field. The images taken during the same orbit are well aligned, and only exhibit marginal, if any, shifts. This enables us to combine the cosmic-ray splits using the `AstroDrizzle` task directly. Most cosmic rays were removed by this step, and the geometric distortions were corrected. However, the pointing of images taken in different orbits can vary by tens of pixels. We selected 2000 bright objects in each of the combined images to align them using the `tweakreg` task. We adopted the first $F555W$ epoch as the reference. The `tweakreg` task updates the WCS information in the image headers, which is then used by the `tweakback` task to update the WCS information of all raw images. Using the updated WCS information, we combined the cosmic-ray splits to obtain 12 aligned, distortion-free frames for each band and each field using the `AstroDrizzle` task. We also obtained master frames by combining multiple images. All the frames were aligned with a precision better than 0.04 pixel.

Table 2.1. Summary of the M101 Observations

Observation	#Epochs	Filter	Camera	Date (MJD)	Obs. Area	Central R.A. (J2000)	Central Dec. (J2000)
06/07F1	12	$F555W$	ACS/WFC	54092.0 - 54121.7	$202'' \times 202''$	14:03:30.2	+54:21:32
	12	$F814W$	ACS/WFC	54092.0 - 54121.7	$202'' \times 202''$	14:03:30.2	+54:21:32
06/07F2	12	$F555W$	ACS/WFC	54094.7 - 54123.8	$202'' \times 202''$	14:02:52.4	+54:20:20
	12	$F814W$	ACS/WFC	54094.7 - 54123.8	$202'' \times 202''$	14:02:52.4	+54:20:20
2013F1	2	$F555W$	WFC3/UVIS	56354.9, 56361.3	$162'' \times 162''$	14:03:31.4	+54:21:12
	2	$F814W$	WFC3/UVIS	56354.9, 56361.3	$162'' \times 162''$	14:03:31.4	+54:21:12
	2	$F160W$	WFC3/IR	56354.9, 56361.3	$162'' \times 159''$	14:03:31.9	+54:21:13
2013F2	2	$F555W$	WFC3/UVIS	56362.9, 56369.6	$162'' \times 162''$	14:02:51.6	+54:20:04
	2	$F814W$	WFC3/UVIS	56362.9, 56369.6	$162'' \times 162''$	14:02:51.6	+54:20:04
	2	$F160W$	WFC3/IR	56362.9, 56369.6	$162'' \times 159''$	14:02:51.8	+54:20:08

The images taken by WFC3 (2013 observations) were pre-reduced by the On-The-Fly-Reprocessing `calwf3` procedure, which performed a basic reduction similar to `calacs` but without CTE correction (Rajan & et al., 2011). We did not perform CTE correction in the reduction, though some columns of WFC3 chips suffered noticable inefficient charge transfering. We reduced the WFC3 images similarly to the ACS data but without cosmic ray rejection.

2.1.3 Photometry

We performed Point Spread Function (PSF) photometry on the images using `DAOPHOT` and `ALLFRAME` (Stetson, 1994). Due to the fact that both two fields are extremely crowded, we derived PSFs using `DAOPHOT` on drizzled images that were generated with TinyTim (Krist et al., 2011). The images contained a grid of 40×40 equally spaced artificial stars. We used the same photometry and PSF-fitting parameters (apertures, sky annuli, PSF fitting radius, and maximum radial extent) for this procedure and our analysis of the M101 data. We generated an initial fiducial object list for each field by performing `DAOPHOT` PSF photometry on the both master frames and their residuals with 3σ detection threshold. The lists were then updated by the `ALLSTAR` procedure, which fit the PSF at all the listed positions in the master frames and removed those that did not meet the fitting criteria. The updated fiducial object lists were used by `ALLFRAME` to simultaneously perform PSF photometry on all the frames. Approximately 500,000 objects were photometered in each field. We carefully selected approximately 50 isolated bright stars in each field to serve as "local standards" and determine frame-to-frame magnitude zero points.

2.2 Cepheid Search Using Template Fitting

We obtained light curves of high variability objects and searched for Cepheids using the `TRIAL` program kindly provided by P. Stetson. We used the Welch-Stetson variability index L (Stetson, 1996) for the initial cut on the Cepheid candidates. We set a threshold

Table 2.2. Magnitude Calibration of M101 Photometry

Band	$m_{\rm psf} - m_{\rm 10pix}$	$m_{\rm 10pix} - m_{\rm inf}$	$m_{\rm inf} - m_{\rm vegazp}$
$F555W$	0.096±0.003	0.092±0.001	−0.724
$F814W$	0.085±0.004	0.087±0.001	−0.501

of $L \geq 0.65$ and further restricted the sample to be brighter than 27 mag in $F555W$ or 26 mag in $F814W$ (instrumental magnitudes).

Due to the relatively small number of epochs, we fit period-dependent Cepheid model light curves (Yoachim et al., 2009) to the measurements with trial periods ranging from 3 days to 100 days. For each object, we fit the model with 1000 log-spaced trial periods and visually checked six locally converged fits with the lowest values of χ^2. We identified 338 and 230 Cepheid candidates in ACS-F1 and ACS-F2, respectively. Figure 2.2 shows six example light curves of the Cepheid candidates.

We calibrated the VI magnitudes of the Cepheid candidates using the transformations derived by Sirianni et al. (2005). The differences between PSF and 10-pixel aperture magnitudes were derived using local standards. The zero points of the calibration are shown in Table 2.2. We applied the transformation from Table 22 in Sirianni et al. (2005) to obtain the Johnson V and Kron-Cousins I magnitudes.

2.3 The Cepheid PLRs in M101

We used reddening-corrected PLRs, including the so-called "Wesenheit index", derived by Ngeow et al. (2009) from OGLE-III observations of LMC Cepheids. The Wesenheit index, defined as $W = I - 1.45 \times (V - I)$, is independent of reddening (Madore, 1982). Figure 2.3 shows the PLRs of the Cepheid candidates in M101. We excluded any objects with color excess $E(V - I)$ less than -0.4 mag or greater than 0.75 mag. We

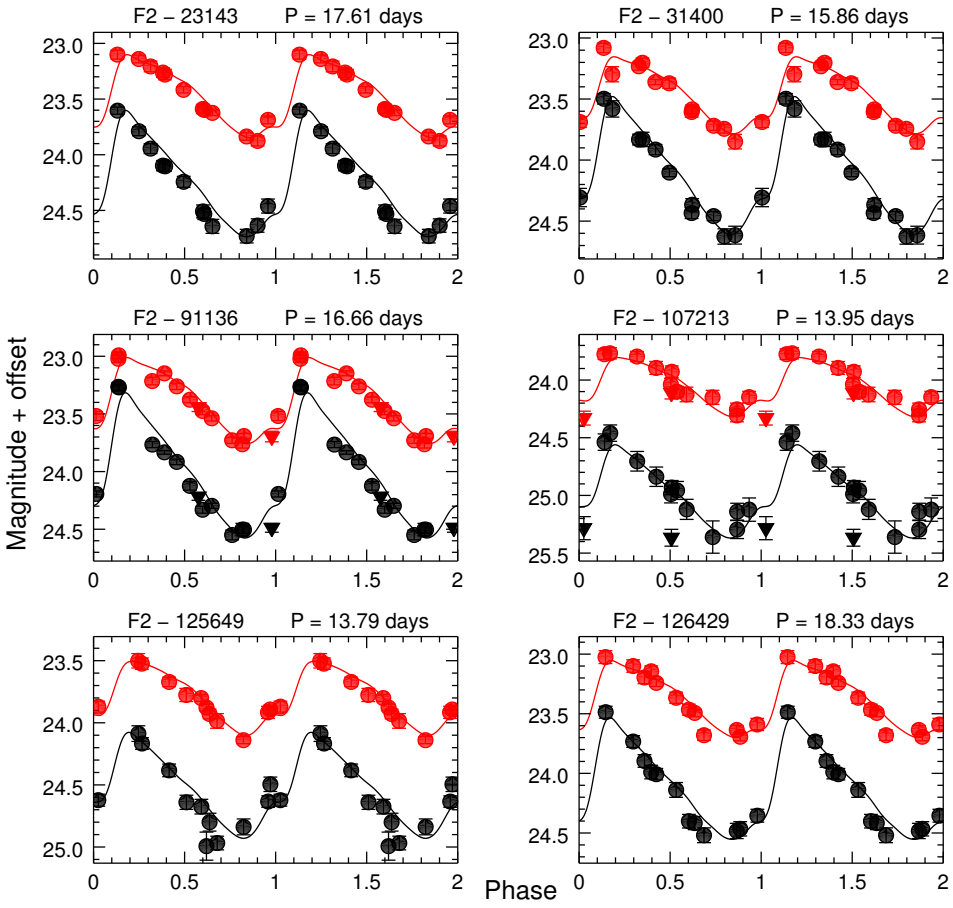

Figure 2.2: Six typical light curves for Cepheid candidates in Field 2. The top (red) and bottom (black) curves represent instrumental $F814W$ and $F555W$ measurements, respectively. Circles and triangles indicate ACS and WFC3 photometry, respectively. The solid lines represent the best-fit templates using the model from Yoachim et al. (2009).

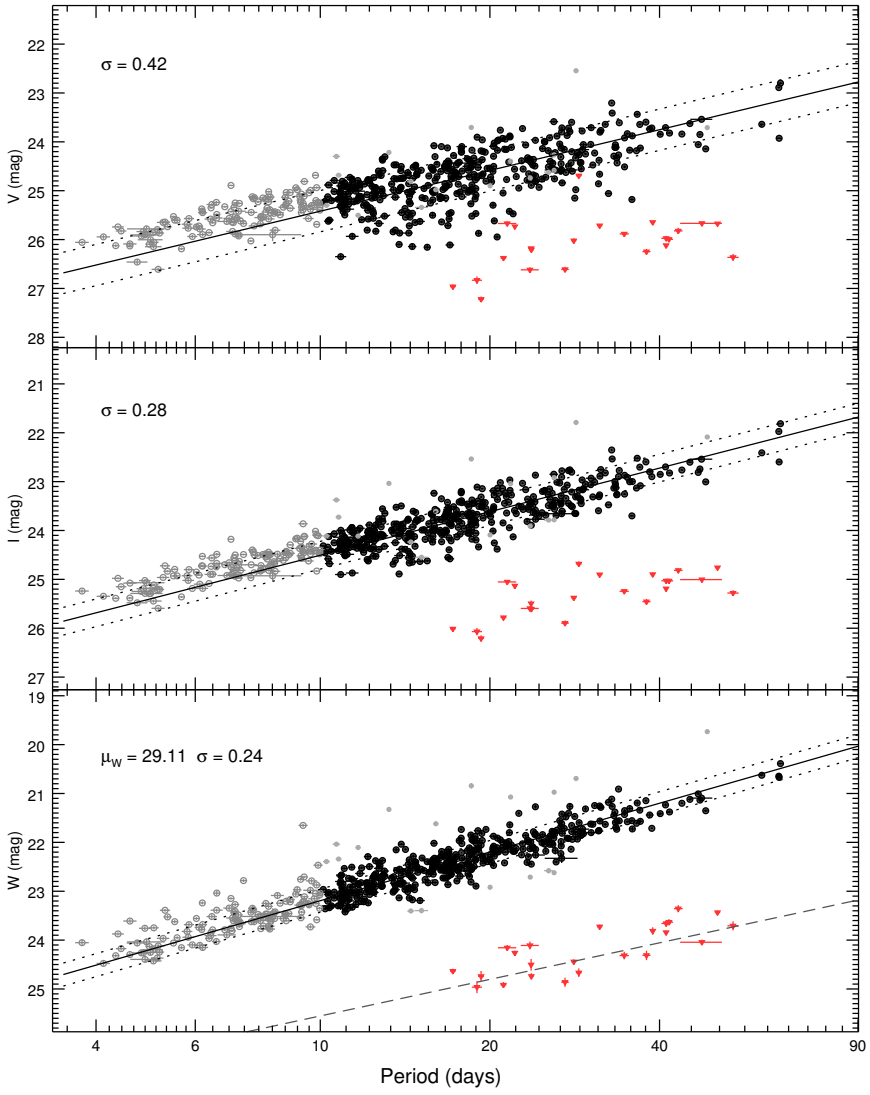

Figure 2.3: PLRs of the M101 Cepheid candidates. We excluded any objects with periods shorter than 10 days (grey open circles). Outliers (grey points) were rejected based on the Wesenheit PLR using an iterative 2.5σ clipping. The red points indicate possible Pop-II Cepheids discovered in this study.

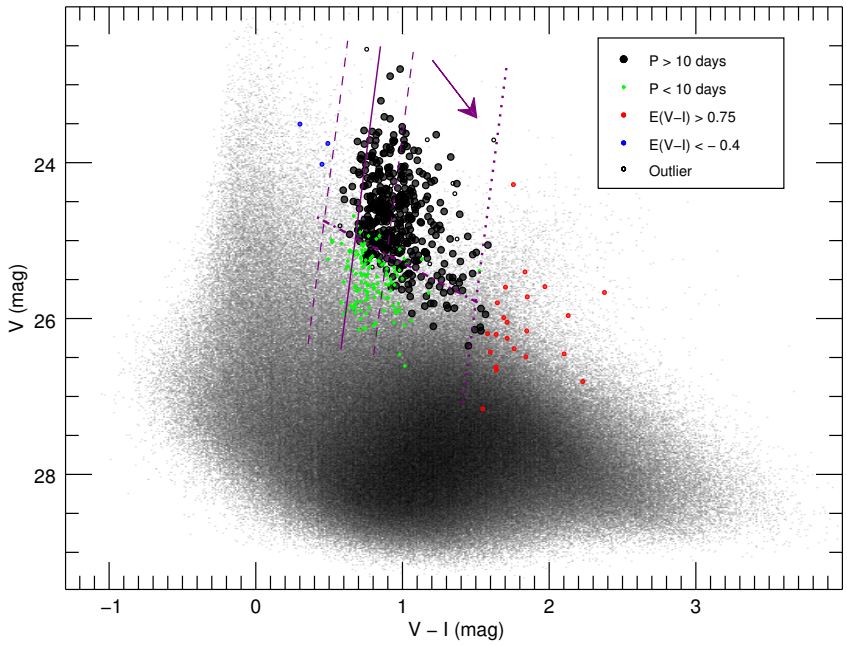

Figure 2.4: Color-magnitude diagram of the Cepheid sample and field stars in M101. Blue open circles indicate PLR outliers. Objects with $E(V-I) < -0.4$ mag and $E(V-I) > 0.75$ mag are indicated by blue points and red points, respectively. The symbol sizes are propotional to the periods of Cepheid candidates. The background grey dots are field stars.

further rejected any objects with periods shorter than 10 days and outliers based on the Wesenheit PLR using an iterative 2.5σ clipping. Figure 2.4 shows the locations of Cepheid candidates and excluded objects in a color-magnitude diagram. We obtained a distance of 29.11 ± 0.08 mag, which is consistent with the distance obtained by NIR observations (Riess et al., 2016, 29.14 ± 0.05 mag).

We tested the metallicity dependence of the Cepheid PLRs in M101. The [O/H] abundances of individual Cepheid candidates were estimated using Equation 5 in Bresolin (2007). We did not find any strong evidence of metallicity dependence, as shown in Figure 2.5.

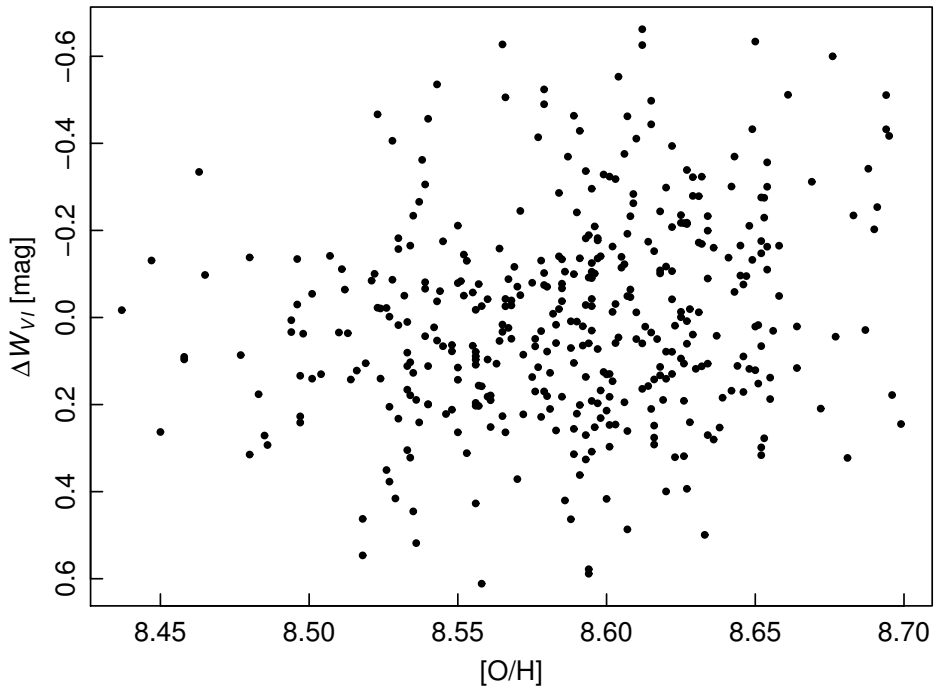

Figure 2.5: Residuals of the Wesenheit PLR against [O/H] abundances for the Cepheid candidates in M101. We did not find any evidence of metallicity dependence.

3. NEAR-INFRARED LIGHT CURVES AND PHASE DETERMINATION FOR 34 GALACTIC CEPHEIDS

As we endeavor to reach a total uncertainty in the Hubble constant of 1% and to search for physics beyond ΛCDM (Riess et al., 2016), it is critical to control and reduce all sources of statistical and systematic error. In particular, a Galactic Period-Luminosity Relation (PLR) based on *Gaia* parallaxes and ground-based H magnitudes would be unacceptably limited to at least 2% systematic uncertainty due to temporal variations in the edges of this bandpass arising from changes in the atmospheric water vapor and OH content. For this reason, starting in *HST* Cycle 19 we began efforts to obtain random-phase photometry of long-period Galactic Cepheids ($P > 7.5$ days) with the same telescope (*HST*), instrument (WFC3-IR) and filters (*F555W, F814W, & F160W*) used to study extragalactic Cepheids. In order to realize the full potential of the H PLR (intrinsic scatter of 0.09 mag, Macri et al., 2015), the space-based random-phase measurements must be corrected to "mean light" values. This requires a firm determination of light-curve amplitudes in this bandpass, as well as precise periods and phase information. These parameters can be obtained in a straightforward manner using ground-based observations, as presented in this section.

3.1 Observations, Data Reduction, and Calibration

Our targets, listed in Table 3.1, are 34 Galactic Cepheids drawn from a parent sample that meets the following selection criteria: (i) $P > 7.5$ d (to complement the mostly short-period samples of Benedict et al., 2007; van Leeuwen et al., 2007); (ii) $E(B–V) < 0.5$ mag (to limit systematic uncertainties from extinction corrections); (iii) $V < 10$ mag and (iv) $D \lesssim 4$ or 5 kpc (to ensure the highest-quality parallaxes from *HST* and *Gaia*, respectively). We focused on targets with $\delta < -20^\circ$ since variables in the parent sample that are located

further north already have suitable light curves from Monson & Pierce (2011). Three objects in common with that publication (T Mon, WZ Sgr & YZ Sgr) were selected to provide a cross-check of our methods.

We refer the interested reader to Casertano et al. (2016) for a detailed discussion of the selection criteria for the *HST* parallax sample (identified by "H" in column 4 of Table 3.1) and a complete list of northern and southern targets. The southern *Gaia* parallax sample (identified by "G" in column 4 of Table 3.1) constitutes a random drawing of suitable objects given that the *HST* observations were scheduled in "snapshot" mode. All *HST* observations for the Cepheids presented here (regardless of the source of their parallax) were executed as part of *HST* GO programs #12679, 12879, 13334, 13334, 13335, 13678, 13686, 13928, 14062 & 14206.

Table 3.1: Galactic Cepheids Target List

Object	R.A. (J2000)	Dec.	Note	Object	R.A. (J2000)	Dec.	Note
β Dor	05:33:38	-62:29:23	C,G	SV Vel	10:44:56	-56:17:24	B,G
T Mon	06:25:13	+07:05:09	B,G	XX Car	10:57:09	-65:08:06	A,G
SS CMa	07:26:07	-25:15:26	A,H	U Car	10:57:48	-59:43:54	B,G
X Pup	07:32:47	-20:54:36	B,H	XY Car	11:02:16	-64:15:46	B,H
VZ Pup	07:38:35	-28:30:00	A,G	XZ Car	11:04:13	-60:58:48	B,H
WX Pup	07:41:59	-25:52:36	A,G	KK Cen	11:42:48	-58:59:36	A,G
AD Pup	07:48:04	-25:34:42	A,G	UU Mus	11:52:18	-65:24:18	A,G
AQ Pup	07:58:22	-29:07:48	B,H	VW Cen	13:33:59	-64:03:18	A,G
BN Pup	08:06:21	-30:05:48	A,G	KN Cen	13:36:37	-64:33:30	B,H
DR Vel	09:31:41	-49:39:18	B,G	V339 Cen	14:21:49	-61:33:00	B,G
l Car	09:45:15	-62:30:28	C,G	SY Nor	15:54:43	-54:34:00	B,G
RY Vel	10:20:41	-55:19:18	B,G	S Nor	16:18:52	-57:54:00	B,G
AQ Car	10:21:23	-61:04:27	A,H	V340 Ara	16:45:19	-51:20:36	A,G
YZ Car	10:28:17	-59:21:00	B,H	RY Sco	17:50:52	-33:42:18	B,H
CR Car	10:32:55	-58:31:18	A,G	W Sgr	18:05:01	-29:34:48	C,G
HW Car	10:39:20	-61:09:09	A,H	WZ Sgr	18:17:00	-19:04:30	B,H
VY Car	10:44:33	-57:33:54	B,H	YZ Sgr	18:49:29	-16:43:24	B,G

3.1.1 Observations

All observations were obtained at the Cerro Tololo Inter-American Observatory (CTIO) SMARTS 1.3-m telescope using the ANDICAM instrument (DePoy et al., 2003) in queue mode. Each variable was observed on 8 -10 epochs, chosen to provide sparse but well-spaced phase coverage (median $\Delta\phi = 0.12$) . For the brightest targets ($H < 6.5$ mag), we used a neutral-density filter that provided $\sim 10^4$ attenuation (hereafter, *ND4*). In these cases, given the large flux difference between the Cepheid and local reference stars, the latter were observed immediately afterwards using only the H filter.

We separated the targets into three groups based on their apparent magnitude and carried out different observation strategies as detailed below. In all cases, each field of view was imaged following a 7-point dither pattern with $\sim 20''$ spacing between adjoining positions, and the total integration time per position was set to ensure we collected at least 10^4 ADU per object (corresponding to signal-to-noise ratio *(SNR)* ~ 100). The Cepheids in group A are faint enough that they do not saturate the detector in the shortest integration time, so we obtained 3×4 s coadds in H at each dither position. The Cepheids in group B are bright enough to saturate the detector in the shortest integration time, so we first used the *H+ND4* filters and then repeated the pattern in H to obtain sufficient *SNR* for the local reference stars. Figure 3.1 gives an example showing the images of the same Cepheid field with and without the *ND4* filter. The Cepheids in group C are so bright that imaging the field in H would lead to considerable persistence artifacts in the detector, so we only acquired data in *H+ND4* and visited a nearby field immediately afterwards to image local standards in H.

3.1.2 Data Reduction and Photometry

We performed the standard corrections (bad-pixel masking, dark current subtraction and flat-field correction) using the IRAF task `ccdproc`. The bad-pixel masks were based

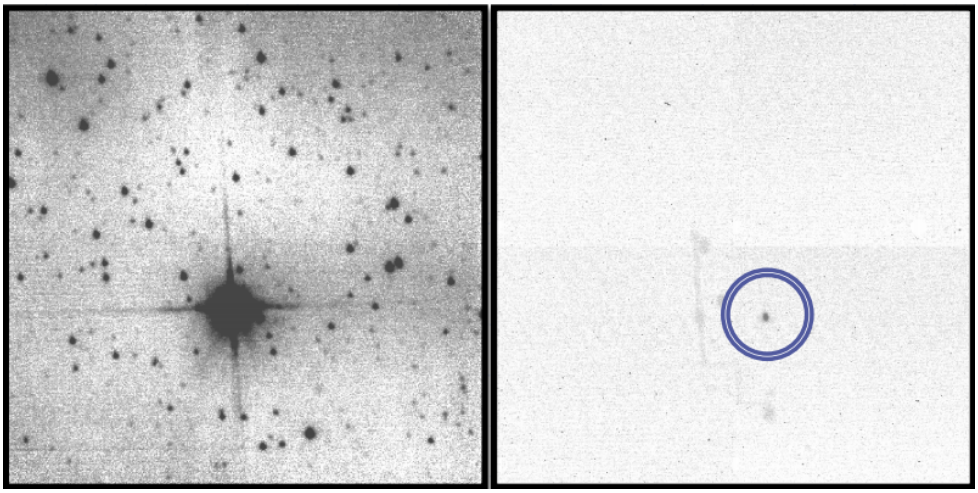

Figure 3.1: Comparison of *H* image (left) and *H+ND4* image (right) of the same field. In the *H* image, the Cepheid (brightest star) is saturated. In the *H+ND4* image, the Cepheid (indicated by a circle) is not saturated but no reference stars are available to derive frame-to-frame zeropoints. Slewing artifacts can be seen in the left and below the Cepheid in the right panel, and they appear in almost all of the *H+ND4* images.

on median-combined dark frames; we additionally masked several regions close to the individual detector boundaries. Bad pixels in the science images were replaced by the average of the highest and lowest valid neighboring pixels using the IDL routine `FIXPIX`. We generated flat-field images based on dome flat observations acquired in H, as the count rate for *H+ND4* was too low to be practical or useful.

We performed aperture photometry on all H frames using the IDL implementation of the `DAOPHOT` routine `APER` and derived geometric transformations using `DAOMATCH` and `DAOMASTER`. We visually inspected all the corrected individual frames and selected $30 - 60$ of those obtained in H with the highest image quality and depth to generate master images using `MONTAGE2` (Stetson, 1993). We generated reference star lists based on these master images, and propagated their positions to the coordinate system of each individual frame. Lastly, we carried out aperture photometry at the predicted position of each star in every frame using the IDL routine `APER`. We excluded any measurements with

photometric uncertainties greater than 0.11 mag.

In the case of the *H+ND4* images, only the Cepheids were visible given the significant attenuation. We visually inspected all images and rejected any that were compromised by slewing artifacts. We carried out aperture photometry in the manner described above.

3.1.3 Magnitude Calibration

We calibrated the master frame photometry using stars in common with the 2MASS Point Source Catalog (Skrutskie et al., 2006), selecting only those objects with good values for the `rd_flg`, `ph_qual`, `cc_flg`, and `bl_flg` flags as explained in the User's Guide to the 2MASS All-Sky Data Release. The median uncertainty in the photometric zeropoint of a field, based on an average of 21 stars per field, was $\sigma_{\rm zp} = 0.013$ mag.

We derived epoch-to-epoch zeropoint offsets for all images of a given field using an average of 10 stars, excluding any that were too bright (saturated or close to the onset of non-linearity), too faint to be useful, located too close to the Cepheid ($r \lesssim 12.5''$), or lying in a small area of the detector with a large fraction of nonlinear pixels. We calculated the error-weighted mean of any given star across the 7 dither positions obtained at each epoch, applying an iterative 3σ outlier rejection. This resulted in 6% of the measurements being discarded.

We derived the *ND4* attenuation factor based on nearly-instantaneous H and *H+ND4* photometry (typical $\Delta t \sim 641$s) of all Cepheids in group B and 5 Cepheids in group A that have additional *H+ND4* observations. Figure 3.2 shows the individual attenuation measurements as a function of instrumental H magnitude, defined as $m = -2.5\log_{10}(\text{cts s}^{-1}) + 25$; the non-linearity in the unattenuated images of targets with $m < 11$ mag (grey points) is evident. We obtained an attenuation factor of 9.324 ± 0.038 mag based on the error-weighted mean of all measurements fainter than that value (black points).

We fit our measurements to the template light curves of Inno et al. (2015). They used

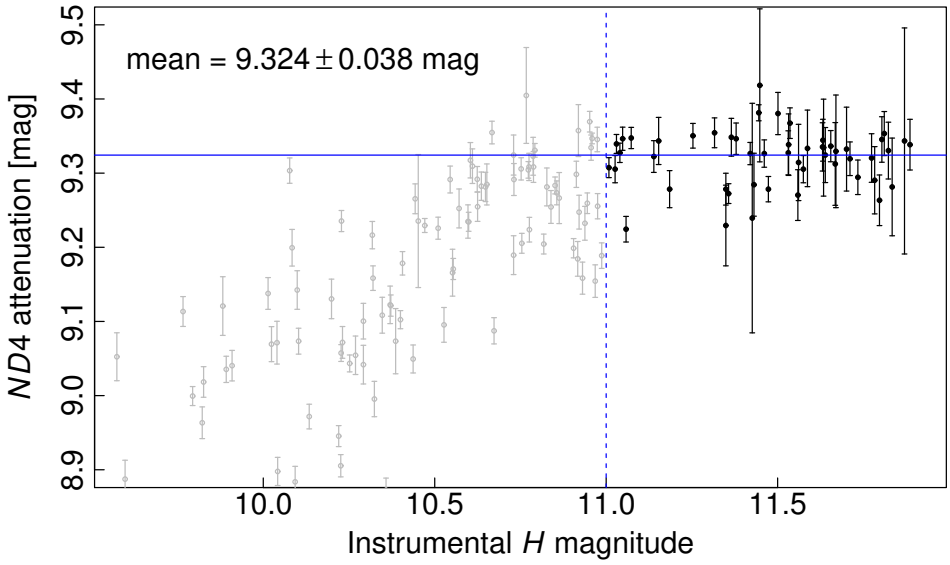

Figure 3.2: The *ND4* attenuation factor versus instrumental H magnitude. The dashed line indicates the approximate onset of non-linearity. The solid line shows the error-weighted mean value of the *ND4* attenuation factor.

a seventh-order Fourier series model, expressed as

$$m_t = M + L \cdot [A_0 + \sum_{i=1}^{7} A_i \cos(2\pi i(\phi_t + \psi) + \Phi_i)] + \sigma_t \epsilon_t$$

with $0 \leqslant \psi < 1$, $L > 0$ and $\epsilon_t \sim \mathcal{N}(0, 1)$. The free parameters M, L and ψ are the mean magnitude, the light curve amplitude, and the phase offset, respectively; A_0, A_i, and Φ_i are model parameters given in Table 4 of Inno et al. (2015) and m_t, σ_t and ϕ_t are the magnitude measurement, its associated uncertainty and the phase at time t, respectively. We computed the latter by adopting a reference time for each Cepheid, t_0, and calculating $\phi_t = (t - t_0/P) \bmod 1$. We carried out an end-to-end test of our procedures by applying the same analysis to the data from Monson & Pierce (2011) for the three variables in common (T Mon, WZ Sgr, YZ Sgr). We found satisfactory agreement of $\langle \Delta M \rangle = -0.014 \pm 0.030$ mag and $\langle \Delta L \rangle = -0.001 \pm 0.040$ mag, given average

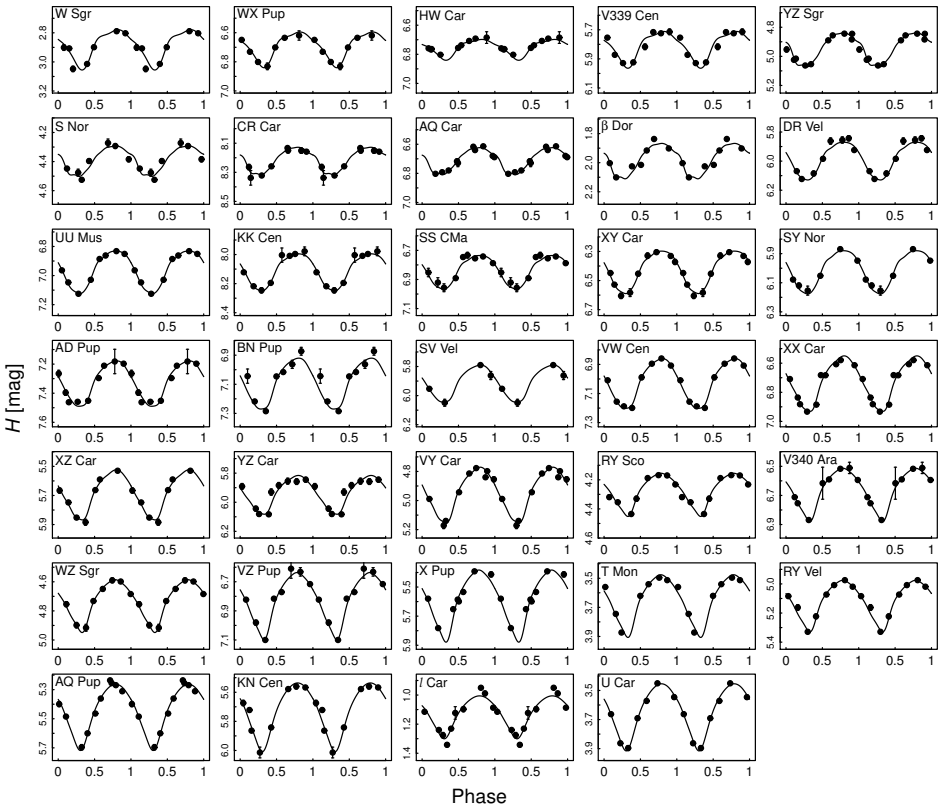

Figure 3.3: H band measurements (points) and best-fit model light curves (solid lines) for all 34 Cepheids. Two cycles of variation are plotted to aid in the visualization of the data.

individual uncertainties of $\sigma_M = 0.04$ mag (fully propagated with σ_{ND4} and $\sigma_{\rm zp}$) and $\sigma_L = 0.025$ mag). Figure 3.3 shows the light curves and best-fit models for all objects in our sample.

3.2 Phase Determination

To improve the phase coverage and constrain the light-curve shapes, we combined our measurements with the H-band observations from Laney & Stobie (1993) for the Cepheids in common. However, due to the long span between the two sets of measurements, inaccuracy or variation in periods can lead to significant phase shifts. To solve for the phase at any given time, we derived the period and period variations using long-span, multi-band

literature measurements.

We based our analysis on five datasets: (1) V-band measurements from the ASAS catalogue of variable stars, (2) V-band measurements from the *INTEGRAL*-Optical monitoring camera, (3) multi-band literature photometry measurements from the McMaster Cepheid Photometry and Radial Velocity Data Archive, (4) JHK measurements from Monson & Pierce (2011) and (5) H-band measurements from this work. Outliers were rejected prior to the analysis.

We fit the combined multi-band measurements with a Fourier series

$$\{1, \cos(k\phi), \sin(k\phi)\} \quad (k = 1, 2, 3, ...)$$

where ϕ is the phase. For each Cepheid, the combined measurements can be described by $\{t, m, \sigma, b\}_{n\times 4}$ where t is time, m is the magnitude, σ is the photometric uncertainty, and b is the band. The dataset is ordered by b. Time t is shifted by $t_{\rm ref}$ and centered at the origin. For each band we constructed a matrix

$$\begin{aligned} B_b = \{&1, \cos(\omega \cdot t_{bi}), \sin(\omega \cdot t_{bi}), \cos(2\omega \cdot t_{bi}), \\ &\sin(2\omega \cdot t_{bi}), ..., \sin(k_b\omega \cdot t_{bi})\}_{n_b\times(2k_b+1)} \end{aligned} \tag{3.1}$$

where t_{bi} is the time of the ith observation in band b, k_b is the order of Fourier decomposition for band b, and $\omega = 2\pi\phi/t$. Since light-curve shape differs as a function of wavelength while the phase does not, we constructed the design matrix as

$$X = \begin{pmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_M \end{pmatrix} \tag{3.2}$$

where M is the total number of bands. All the blocks share the same ω while their Fourier coefficients can be different.

We performed a grid search on ω by least squares. Given ω, the regression coefficients were solved by

$$\beta = (X^T \Sigma^{-1} X)^{-1} (X^T \Sigma^{-1} Y) \tag{3.3}$$

where $\Sigma_{n \times n}$ is a diagonal matrix with $\mathrm{Diag}(\Sigma) = \{\sigma^2\}_n$, $Y = \{m\}_n$. The residual sum of squares (RSS) is subsequently obtained by

$$\mathrm{RSS} = (Y - X\beta)^T \Sigma^{-1} (Y - X\beta) \tag{3.4}$$

We tried a model of constant period with $\omega = 2\pi/q_0$ and a model of time-dependent period with $\omega = 2\pi/(q_0 + q_1 \cdot t)$ for each Cepheid. Due to the fact that photometric uncertainties are not always available and the scale factors of uncertainties can be systematically different across different sources, we used equal weights across the data. We found 18 out of the 34 Cepheids show period variations. For U Car, the period changing rates are only applicable to the date close to reference times. The best-fit parameters for constant-period and time-dependent period variables are shown in Table 3.2 and Table 3.3, respectively. Figure 3.4 shows the phase-folded light curves of AQ Pup as an example.

We compared our determination of periods with the General Catalogue of Variable Stars (GCVS; Samus et al., 2017)*. For the objects in Table 3.2, our results agree with the GCVS periods. Figure 3.5 shows the differences between our determination of periods and the GCVS periods for the 34 Cepheids, with the Cepheids in Table 3.3 being indicated by red points. We adopted the periods at the reference times in Table 3.3 for those Cepheids with changing period.

We found an interesting trend for the Cepheids in our sample; the period-changing

*http://www.sai.msu.su/gcvs/gcvs/

Table 3.2. Phase Parameters for Cephedis with Constant Periods

Object	q_0 [day]	σ	$t_{\rm ref}$ [MJD]
β Dor	9.842852	2.83E-05	48966.1
VZ Pup	23.175399	5.51E-05	48838.1
BN Pup	13.672585	2.59E-05	49461.2
DR Vel	11.199335	3.01E-05	48414.1
AQ Car	9.769438	1.98E-05	49532.7
HW Car	9.199174	1.02E-04	53524.9
XZ Car	16.652481	3.22E-05	48476.9
UU Mus	11.636187	1.23E-05	49414.8
VW Cen	15.037048	5.03E-05	48896.1
V339 Cen	9.466580	2.48E-05	49505.4
SY Nor	12.645328	2.70E-05	47894.4
S Nor	9.754850	5.00E-05	51041.6
RY Sco	20.321080	6.49E-05	49488.0
W Sgr	7.595008	8.82E-05	43670.5
WZ Sgr	21.851189	4.48E-05	49468.6
YZ Sgr	9.553801	2.03E-05	50608.7

Table 3.3. Phase Parameters for Cepheids with Time-Dependent Periods

Object	q_0 [day]	$\sigma(q_0)$	q_1	$\sigma(q_1)$	$t_{\rm ref}$ [MJD]	$\dot{P}$ [second/year]	$\sigma(\dot{P})$
T Mon	27.033647	3.5E-04	5.7123E-07	4.5E-08	50405.8	36.05	2.86
SS CMa	12.355863	2.9E-05	-2.3166E-07	5.1E-09	49153.1	-14.62	0.32
X Pup	25.967493	6.7E-05	4.7212E-07	9.4E-09	48871.9	29.80	0.59
WX Pup	8.936743	4.5E-05	-3.4725E-08	5.5E-09	48679.7	-2.19	0.35
AD Pup	13.596511	3.3E-05	7.3398E-08	5.2E-09	48685.3	4.63	0.33
AQ Pup	30.121556	1.0E-04	4.0285E-06	2.2E-08	49372.1	254.26	1.41
l Car	35.552648	1.0E-03	1.1165E-06	4.8E-07	50903.6	70.47	30.29
RY Vel	28.131528	1.3E-04	5.9027E-07	2.5E-08	48910.8	37.25	1.59
YZ Car	18.166716	3.7E-05	1.2494E-07	8.3E-09	48440.3	7.89	0.52
CR Car	9.759351	4.8E-04	-8.5781E-08	4.7E-08	47919.8	-5.41	2.98
VY Car	18.898434	2.5E-04	-1.3034E-06	2.6E-08	49416.6	-82.27	1.64
SV Vel	14.097462	2.3E-05	4.6633E-08	4.3E-09	48430.9	2.94	0.27
XX Car	15.708239	3.5E-05	-1.4270E-07	4.9E-09	48350.2	-9.01	0.31
U Car	38.837711	3.5E-04	2.3758E-06	6.1E-08	49116.0	149.95	3.88
XY Car	12.435852	2.0E-05	4.6595E-08	3.6E-09	48351.6	2.94	0.22
KK Cen	12.182007	5.5E-05	5.5715E-08	1.1E-08	48411.3	3.52	0.69
KN Cen	34.033685	3.1E-04	-1.1125E-06	2.8E-08	47506.0	-70.22	1.78
V340 Ara	20.810239	6.2E-05	3.3876E-07	1.4E-08	48376.4	21.38	0.86

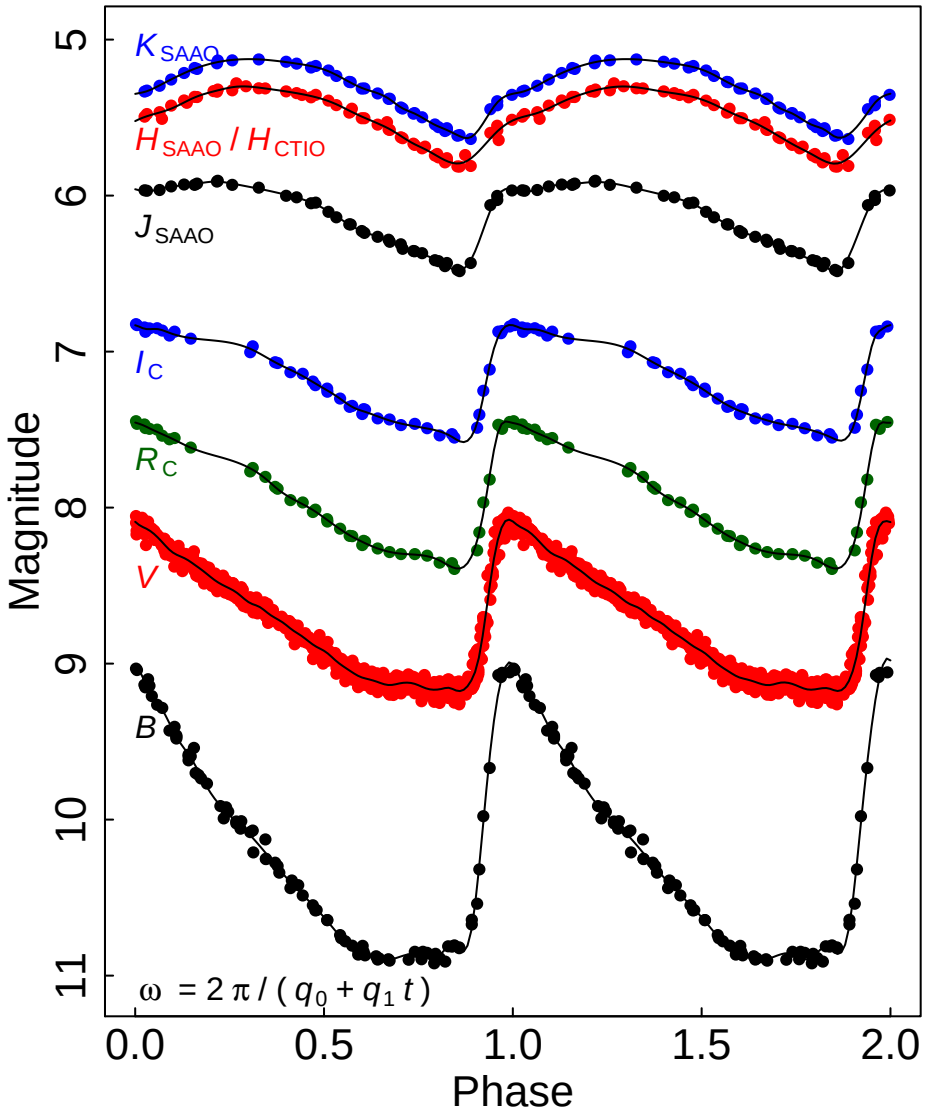

Figure 3.4: Phase-folded light curves for AQ Pup with time-dependent period. The time span of the measurements is more than 43 years.

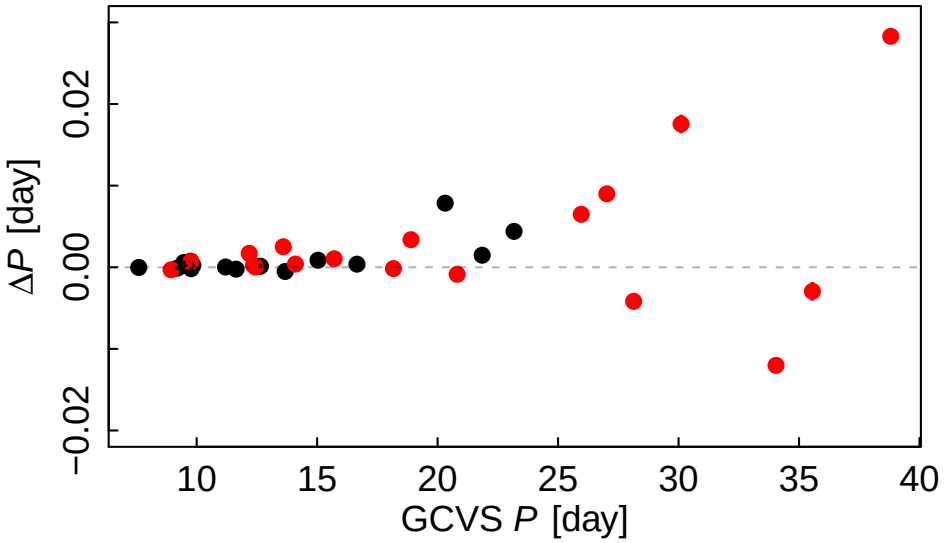

Figure 3.5: Differences in periods between our determination and the GCVS values. The black points indicate the Cepheids with constant periods while the red points indicate the Cepheids with changing periods. The gray dashed line indicates the location of zero difference.

rates $\dot{P}$ are correlated to the periods P. Figure 3.6 shows $|\dot{P}|$ against P for the Cepheids in Table 3.3, together with 3 more Galactic Cepheids (RW Cas, S Vul, and VX Per) whose parameters were derived using the same method. We fit a straight line to the logarithm of $|\dot{P}|$ versus P and obtained the empirical relation

$$\log_{10}(|\dot{P}|) = -0.007_{\pm 0.095} + 0.058_{\pm 0.004} \cdot P \qquad (\sigma = 0.22). \tag{3.5}$$

Similar relations have been studied by Fernie (1984), who suggested the period changes are due to stellar evolution.

3.3 Results

We tested the accuracy of our phase-correction method with simulations. We down-sampled the well-observed Large Magellanic Cloud (LMC) Cepheids from Persson et al.

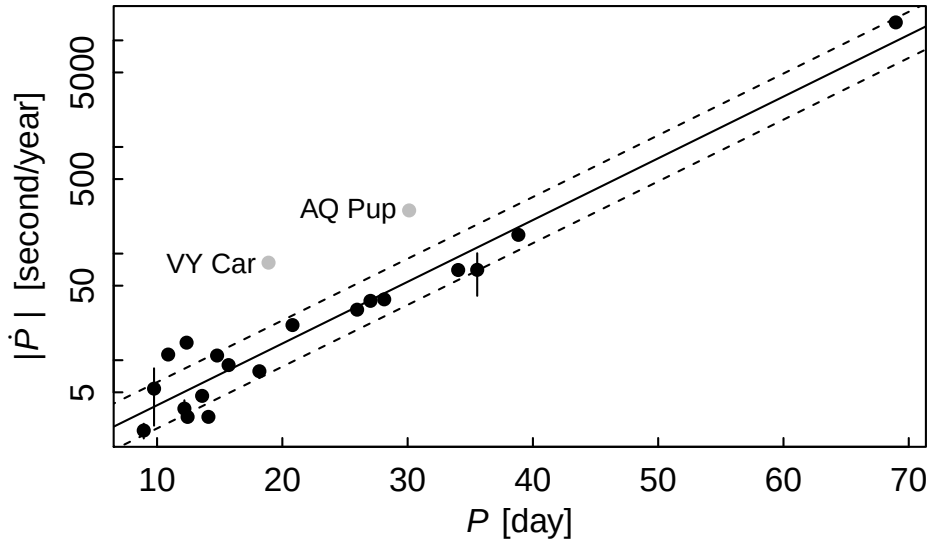

Figure 3.6: Correlation between the absolute period-changing rates and periods for 21 Galactic Cepheids. Two Cepheids (VY Car and AQ Pup, grey points) were excluded in the fit. The solid line is the best-fit result while the dashed lines indicate the 1σ scatter.

(2004) to match the cadence and photometric uncertainties of our Galactic Cepheid observations. To do this, we firstly retrieved their *H* measurements and cross matched them to the OGLE-III Cepheid catalog, and rejected any objects with less than 20 *H* measurements. We then fit the Inno et al. (2015) model to the well-sampled LMC Cepheid measurements in order to derive continuous functions. We down-sampled the continuous functions at the same phases of the Galactic Cepheid observations in our sample, but with overall phase shifts. Photometric noise was finally added to the down-sampled light curves based on the CTIO observations.

We applied the template-fit method to these noisy down-sampled light curves and computed their mean magnitudes and magnitudes at the *HST* epochs. The differences $H_{\rm corr}$ between mean magnitudes and magnitudes at *HST* epochs were then compared to the values computed from the continuous light curves, which were considered as the "true values".

For the Cepheids with multiple *HST* epochs, we averaged $\Delta H_{\rm corr}$ for all the *HST* epochs to obtain the overall accuracy. Figure 3.7 shows the differences in $H_{\rm corr}$ between the results of down-sampled light curves and the "true values" for the simulation of 29 Galactic Cepheids in our sample (the rest of them do not have LMC counterparts with similar periods). We found an overall accuracy of 0.005 mag for the correction.

Figure 3.8 shows the light curve and best-fit model for XY Car, along with the *HST* observations (Adam Riess, private communication). XY Car is one of the objects with mulitple *HST* measurements to check the ground photometry. We found good agreement between our ground-based and the *HST* measurements.

3.4 Summary

We obtained ground-based H-band light curves for 34 nearby ($D \lesssim 5$ kpc) Galactic Cepheids recently observed with *HST*. We redetermined their periods and derived the associated uncertainties. We fitted our measurements using templates and derived light curve amplitudes and contemporary phase information. We tested our correction method with simulations and derived an accuracy of 0.005 mag. Our results will be used to correct random-phase *HST* observations to mean light.

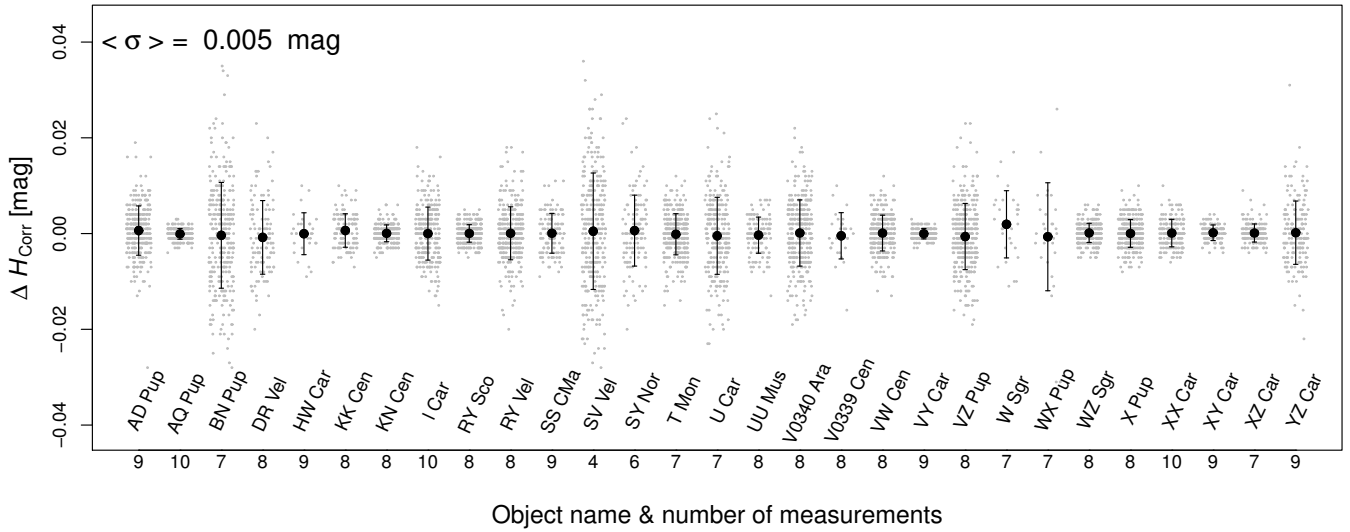

Figure 3.7: Differences in magnitude between the corrections from simulated Cepheid light curves and the "true values" for most objects in our sample. The number below the horizontal axis is the number of ground measurements. Five variables in the sample are missing due to the lack of counterparts in the LMC with similar periods.

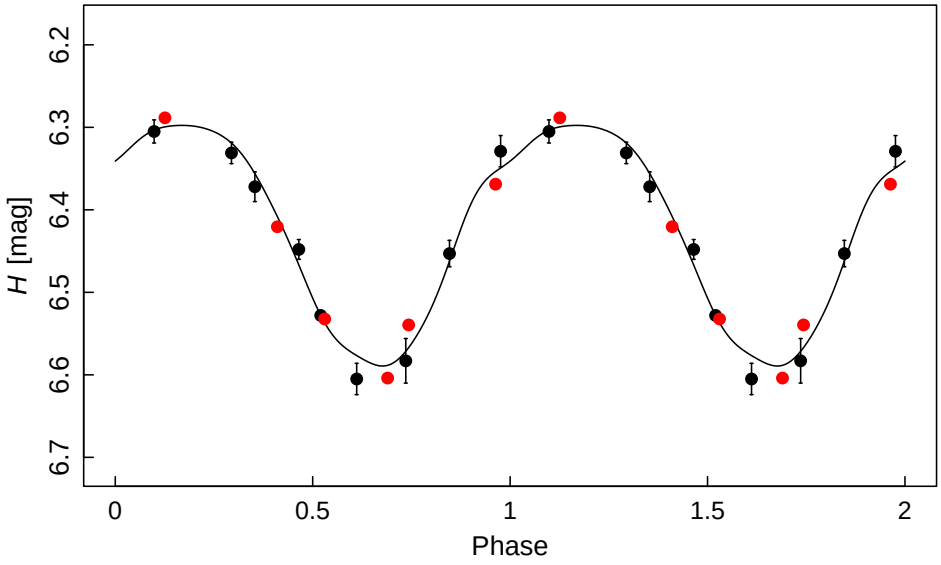

Figure 3.8: Ground-based light curve (black points) and best-fit model (black curve) for XY Car. Red points indicate measurements obtained with *HST*. A zeropoint offset was applied to the *HST* measurement to account for the difference in filters.

4. MIRA SEARCH IN M33*

In this section, we report the results of a search for Mira candidates in M33 using I-band observations spanning nearly a decade that were obtained by the DIRECT project (Macri et al., 2001, hereafter M01) and by Pellerin & Macri (2011, hereafter PM11). Traditional periodogram methods, such as Lomb–Scargle (Lomb, 1976; Scargle, 1982), are not optimal for this search due to relatively sparse temporal sampling, large gaps between observing seasons, and the expected long-term variations in Mira light curves. We developed a novel semi-parametric periodogram technique (He et al., 2016, hereafter H16) based on the Gaussian process method that contains a data-driven component in the model light curve to account for deviations from strict periodicity and gives an overall better performance. We coupled this algorithm to random forest (RF) classifiers, training and testing them extensively on simulated light curves.

4.1 Observations and Data Reduction

We based our search on the observations of M33 obtained by M01 and PM11. These surveys covered most of the disk of this galaxy with a combined baseline of nearly a decade (1996 September to 1999 November for M01; 2002 August to 2006 August for PM11) mainly using the Fred L. Whipple Observatory (FLWO) 1.2 m and the WIYN observatory 3.5 m telescopes with a variety of cameras (see the respective publications for details). While images were obtained in multiple bandpasses (BVI), our analysis is only based on the I-band time-series photometry because Mira candidates fall below the detection limit in the bluer bands. Given that both studies had to rely on multiple pointings to cover the area of interest and not all locations were observed on a given night, the sampling pattern

*Material in this section has been published in "The M33 Synoptic Stellar Survey. II. Mira Variables" by Yuan et al. (2017), *The Astronomical Journal*, 153:170 (11pp). Copyright 2017 by the American Astronimical Society.

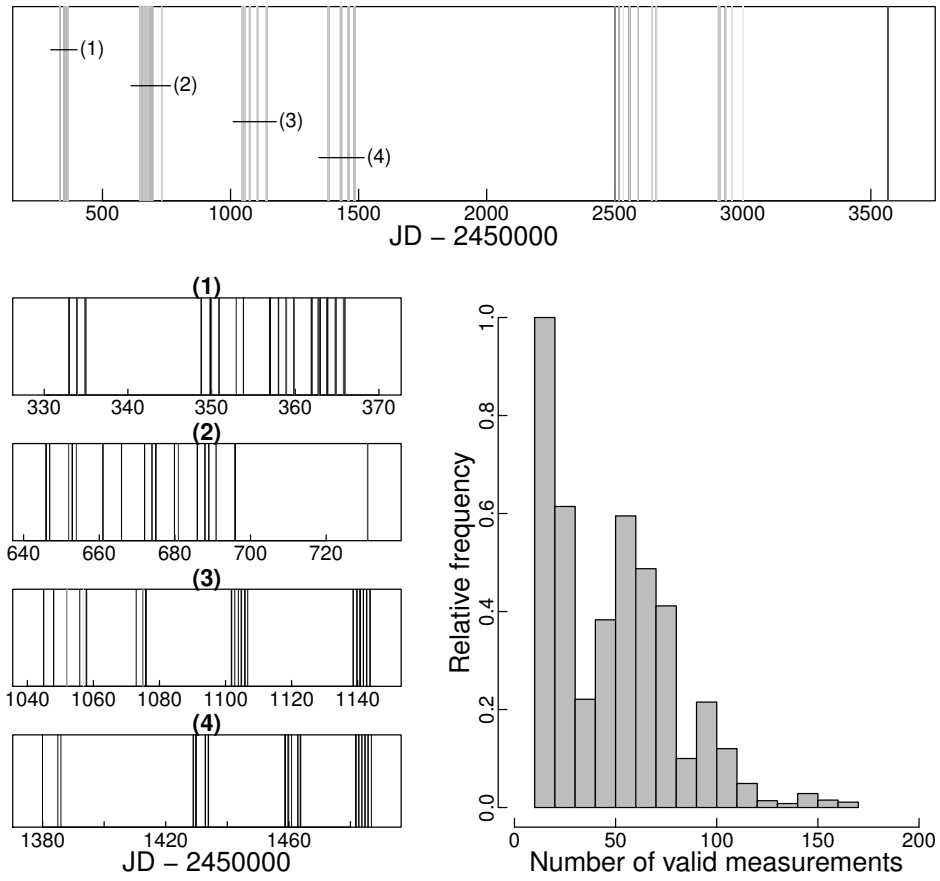

Figure 4.1: Top: cadence of M33 observations in I by M01 and PM11. The grayscale levels are linearly proportional to the number of measurements per square arcminute of each epoch. Bottom left: expanded view of the cadence for seasons 1–4. Bottom right: histogram of measurements for stars with N > 10 and $I < 21.45$ mag.

varies considerably across the disk. Figure 4.1 shows the overall sequence of observations, of which only a subset will be applicable at a given position.

We performed new photometric measurements on the preprocessed images from PM11 to mitigate issues arising from geometric distortions and poor image registration at the corners of each field (which corresponds to a single telescope and camera). Unlike the approach of the previous work, we first analyzed the images of a given field and later

combined the photometry for matching sources. We obtained aperture and PSF photometry using the DAOPHOT, ALLSTAR, and ALLFRAME programs (Stetson, 1987, 1994) in a quasi-automatic manner by integrating the tasks into an R script pipeline. We defined the PSF for each image using the top 50 bright and isolated stars and selected the one with the sharpest PSF from each field to serve as a reference for ALLFRAME. We selected a larger number of secondary standards for image registration and to tie the photometric measurements. These were among the brightest few percent of all sources in a given field and had photometric uncertainties below 0.02 mag. We determined frame-to-frame zero-point offsets, computed mean instrumental magnitudes, and extracted light curves using TRIAL (Stetson, 1996).

We obtained astrometric and photometric calibrations for each field using the catalog published by the Local Group Galaxies Survey (Massey et al., 2006, LGGS). We derived the astrometric solution for the reference frame of each field using WCSTools (Mink, 1999). We matched the LGGS sources to the star list from the (now astrometrically calibrated) reference frame of each field and solved the transformation equation

$$I_c = (1 + a) \cdot I_i + b\,, \tag{4.1}$$

where I_c is the magnitude in the standard system (Kron–Cousins I for LGGS), I_i is the instrumental PSF magnitude of the reference frame of a given field, b is the zero-point offset, and a provides a simple correction for color terms and/or photometric biases due to crowding (given the considerable variation in stellar density across the disk and in image quality among the fields). We were not able to apply a traditional photometric transformation with zero-point and color terms because we only have single-band (I) photometry for the vast majority of the sources. We solved for the coefficients using the top 25% and 10% brightest stars in the fields imaged at WIYN and FLWO, respectively, applying an

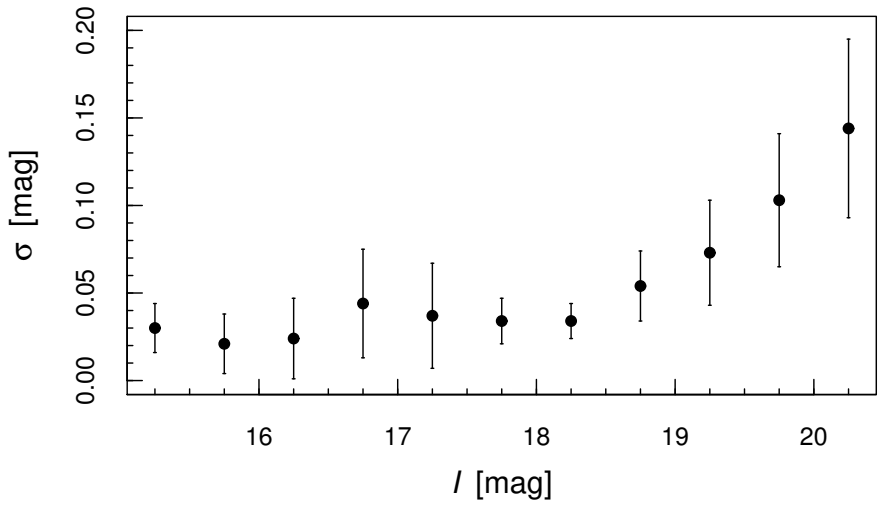

Figure 4.2: Photometric precision for secondary standards as a function of magnitude.

iterative outlier rejection of 3σ and 2.3σ, respectively. The median value of a was 0.001, with 95% of the values falling between -0.024 and +0.015.

Given the significant overlap between the fields of M01 and PM11, most objects have multiple light-curve segments that were merged as follows. If two sources in different fields had coordinates that matched to better than $1''$, and there were no other sources detected within $1''$, then they were considered the same object. If there were neighboring sources within that radius, then the closest object with a magnitude difference less than 0.5 mag was selected (recall that each field was already transformed to the standard system prior to this step). We ensured that at most one source in one field could merge with one source in another field. We tested the photometric precision of the aforementioned steps by reidentifying the local standards of each field and examining the dispersion of the merged light curves relative to the dispersion of individual segments. Figure 4.2 shows that we reached a photometric precision of 0.03 mag.

We selected 239,907 light curves for the Mira search, rejecting any with fewer than

10 measurements or with mean magnitudes fainter than I = 21.45 mag. The first cut is based on extensive testing via the simulated light curves (Section 4.2) of our algorithm (Section 4.3); the procedure does not yield reliable periods for sparser samplings. The second cut is due to the large photometric uncertainties beyond that magnitude limit, which prevent the detection of light-curve variations of the expected amplitude.

4.2 Simulated M33 Light Curves

We simulated 10^5 light curves of Miras and the same number of semi-regular variables (SRVs) that accurately reproduce the photometric uncertainties and temporal sampling of the M33 data set. The simulated light curves were used to test our period determination algorithm and to train a classifier to identify Mira candidates. The methods used in the simulation are based on the ones we developed for H16 and rely on very high precision I-band light curves sampled at hundreds of epochs over 7.5 yr by phase III of the OGLE project (Udalski et al., 2008). We also generated an equal number of artificial light curves of "constant" stars, in order to properly balance the training data for the classifier.

4.2.1 Miras

The procedure used to fit templates to the OGLE-III Mira light curves is explained in detail in Section 4.3 of H16. Briefly, the light curve is decomposed into a mean value, a regular variation of period P, a low-frequency (long-term) trend, and a high-frequency/small-scale term (see Figure 6 in H16). The latter three components are modeled by a Gaussian process with different kernels. The maximum likelihood method is used to obtain the model parameters as a function of the trial value of P. Once the best-fit model is found, it can be used to predict the magnitude at any time during the observation baseline, including the brightest value reached by the variable (I_m). This quantity is of interest because Kanbur et al. (1997) have suggested that PLRs at maximum light may have a significantly smaller dispersion than at mean light.

Simulated light curves were generated by sampling the best-fit model using randomly selected observing patterns from the actual light curves, with equal probabilities. We shifted the starting point of each simulated light curve by a random value Δt, limited in range only to ensure that the resulting light curve was still contained within the span of the OGLE observations. These random shifts helped to obtain many unique simulated light curves when using the same template. We applied a shift of $\Delta m = +6.2$ mag to account for the approximate difference in apparent I-band distance moduli between the LMC and M33 ($\Delta\mu_0 = 6.26 \pm 0.03$ mag and $\Delta A_I = -0.05$ mag, based on PM11; Schlafly & Finkbeiner, 2011). Furthermore, we introduced a realistic amount of photometric noise following the procedure outlined in Section 4.1 of H16.

As a final step in our simulations, we took into account the fact that the OGLE LMC observations reach substantially deeper in terms of absolute magnitude than the M33 observations and considered the possibility that the light-curve shape of Miras may be a function of their luminosity. If the latter is true, a mismatch of the luminosity functions would bias our classifier. We derived the completeness function of the M33 photometry by fitting the observed luminosity function with an exponential, and we applied it to the luminosity function of the OGLE LMC Mira candidates after offsetting the latter by the difference in apparent distance moduli between the two galaxies. We then randomly selected simulated Mira light curves such that we reproduced the observed luminosity function of the M33 photometry.

Figure 4.3 shows a representative example of a simulated Mira light curve that mimics the cadence and photometric precision of the actual M33 data, while Figure 4.4 shows the completeness function of the M33 photometry.

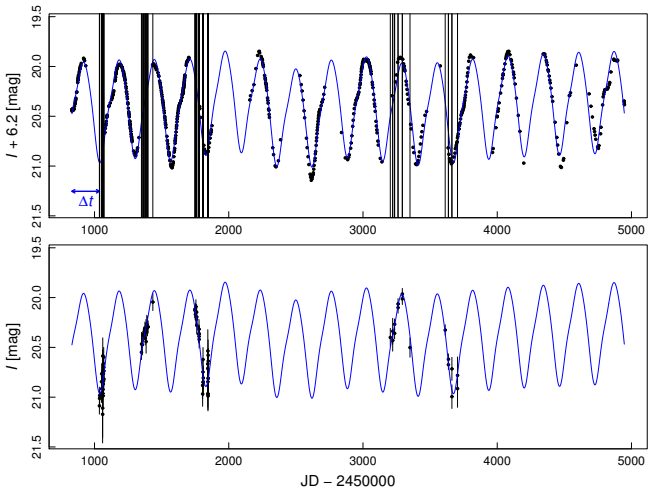

Figure 4.3: Example of a template Mira light curve and simulated M33 measurements. Top: OGLE measurements of a Mira candidate in the LMC (black points), best-fit template using our model (blue curve), and sampling pattern of one of the M33 fields (vertical black lines). The horizontal blue arrow indicates the random shift applied to the pattern to sample the light curve. Bottom: corresponding simulated M33 light curve, including additional photometric noise.

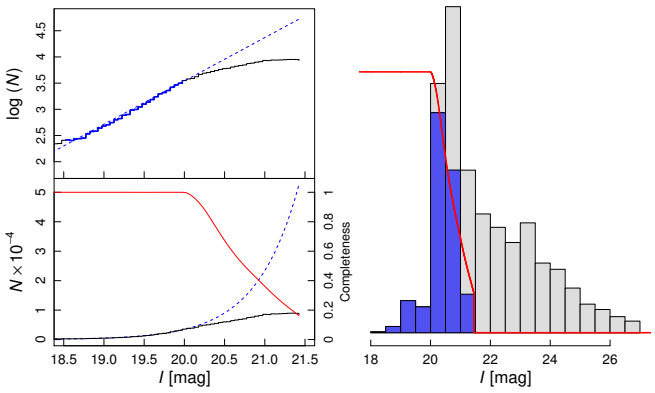

Figure 4.4: Left: derivation of the empirical completeness function for M33 photometry (top: logarithmic; bottom: linear scale). An exponential model is fit to the observed luminosity function (solid black line) over the magnitude range (solid blue line) and extrapolated over the range plotted with a dotted blue line. The derived completeness function (solid red line) is shown in the bottom panel only. Right: magnitude distribution of Mira template light curves before (gray) and after (blue) convolution, with the completeness function (red line).

4.2.2 SRVs and "Constant" Stars

The light curves of SRVs share some similarities with those of Miras (cyclic variations), although they tend to be more chaotic and less periodic and usually exhibit smaller amplitudes. Nevertheless, since they outnumber Miras six to one in the catalog of Soszyński et al. (2009), ignoring them could significantly bias our classifier. Hence, we included simulated SRV light curves in the training data. We obtained templates by applying a smoothing spline to the OGLE observations and generated artificial light curves by following the same procedure as for Miras (sampling based on the M33 observing patterns, random shifts of the starting point, convolution with the M33 completeness function, and addition of photometric noise).

Lastly, in order to balance the various types of objects that are used to train the classifier, we simulated the light curves of "constant" stars by randomly shuffling the observation times of all light curves in our data set while keeping the original magnitudes and uncertainties. The shuffling removes any potential periodicity in the original data and allows the generation of multiple artificial light curves from the same object.

4.3 Semi-Parametric Model for Identification and Period Determination of Miras

Given the stochastic variations exhibited by Mira light curves at optical wavelengths, traditional algorithms become less efficient at discovering these objects and obtaining reliable periods in the limit of the sparsely sampled observations. We showed in H16 that our semi-parametric model provides overall improvement of period recovery in this regime. We applied this model to the M33 observations and coupled it to an Random Forest classifier in order to identify Mira candidates.

We refer interested readers to Section 3 of H16 for a detailed description of our semi-parametric model and its performance, which are only briefly summarized here. The model is based on Gaussian process regression, which has been previously applied to astronomi-

cal time-series observations. For example, Faraway et al. (2016) modeled the light curves of several types of transient event from the Catalina Real-Time Transient Survey with a squared exponential kernel, while Aigrain et al. (2016) applied this technique to *Kepler* data to correct systematic trends in their photometry.

The semi-parametric model we used is a simplified version of the one described in Section 4.2 to account for the quality of the M33 data. Given a set of measurements over n epochs, $\mathcal{D} = \{(t_i, m_i, \sigma_i)\}_{i=1}^{n}$, with t, m, and σ representing time, magnitude and measurement uncertainty respectively, the model is

$$m_i = m + \beta_1 \cos(2\pi f t_i) + \beta_2 \sin(2\pi f t_i) + h(t_i) + \sigma_i \epsilon_i \tag{4.2}$$

where f is the frequency (day^{-1}), $h(t)$ is modeled by a Gaussian process with the squared exponential kernel

$$k(t_1, t_2) = \theta_1^2 \exp\big(-\frac{(t_1 - t_2)^2}{2\theta_2^2}\big),$$

and the amplitude of the periodic component is $A_P = 2(\beta_1^2 + \beta_2^2)^{1/2}$. The parameters m, β_1, and β_2 are assumed to follow Gaussian priors and integrated out of the likelihood function when estimating other parameters. Optimization is performed over hyper-parameters θ_1 and θ_2 for each trial value of f, ranging from 5×10^{-4} to 10^{-2} every 10^{-5}. The log-likelihood function Q (the "frequency spectrum") is evaluated at each trial frequency (see Equation (10) in H16).

We applied the model to all simulated and real light curves and adopted the highest peak in the frequency spectrum (f_1) as the most likely frequency. We found that the true period was successfully recovered (with a tolerance of $|f_1 - f_{\rm true}| < 2.7 \times 10^{-4}$, as defined in H16) for 69.4% of all simulated Mira light curves.

We estimated the period uncertainties for all light curves using a nonparametric boot-

strap approach followed by error scaling, as follows. First, we resampled the measurements with replacement and derived new values of f_1, repeating this procedure 500 times per variable. We used the standard deviation of the results for each object as an initial estimate of the period uncertainty. Next, we carried out the same procedure on the simulated Miras (with 30 iterations per light curve) and calculated $\delta P = (P_i - P_r)/\sigma(P_r)$, where P_i and P_r are the input and recovered periods and $\sigma(P_r)$ is the bootstrap-based uncertainty for the latter. Restricting our analysis to the successfully recovered variables (as defined in the previous paragraph), and under the assumption that period residuals should follow a Gaussian distribution, we calculated the fraction with $|\delta P| < 1$ and iteratively rescaled $\sigma(P_r)$ until 68.3% of the objects met that criteria. This required a rescaling factor of 2.33, which was then applied to the bootstrap-based uncertainty estimates of the Mira candidates.

4.4 Results

4.4.1 Random Forest Classification of Miras

A machine-learning technique, Random Forest (RF), has already proven to be effective in classifying different classes of variable stars (Dubath et al., 2011; Richards et al., 2011). We built an RF classifier based on the model parameters, the features of the frequency spectra, and information obtained from the simulated light curves, as detailed below. Once trained on the simulated data, it was applied to the M33 observations to select Mira candidates. Our choice of RF is supported by a comparative study reported in Section 4.4.2, where it is shown to outperform several state-of-the-art classifiers on simulated data. Figure 4.5 shows the frequency spectra for a representative artificial light curve for each of the three classes (Mira, SRV, and constant star). The frequency spectra of SRVs and constant stars are usually quite different due to their lack of periodicity, which indicates that the shape of this function can be used to identify Mira candidates.

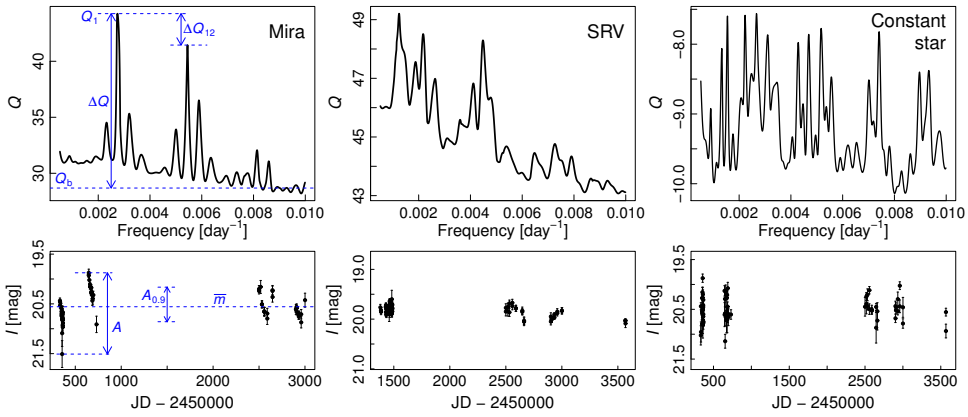

Figure 4.5: Examples of frequency spectra (top) and corresponding light curves (bottom) for a simulated Mira (left), SRV (middle), and constant star (right). The blue dashed lines and arrows indicate some of the quantities used as classification features.

We extracted seven features from the best-fit model parameters, four from the frequency spectra, and seven from the light curves. Table 4.1 provides a summary of all of the features and their rank in terms of importance for separating Mira candidates from other stars and for separating the former into likely C-rich or likely O-rich. The upper left panel of Figure 4.5 shows a graphical representation of the features that were extracted from the frequency spectra. Figure 4.6 shows two features we extracted from the light curves: the standard deviation of the residuals from piecewise quadratic fits ($\sigma(R_q)$) and its ratio to the scatter about the overall mean value ($\sigma(\overline{m})$). This ratio is significantly smaller for Miras than for any of the other classes.

Table 4.1. Features for the Classifier

Feature	Description	Src[a]	Rank[b]	
			Mira	C/O
ΔQ	Difference of log likelihoods $Q_1 - Q_b$ (see below)	F	1	12
$\sigma(R_q)/\sigma(m)$	Ratio of standard deviations (see below)	L	2	9
$A_{0.9}$	Light-curve range from 10th to 90th percentile	L	3	5
ΔQ_{12}	Difference in log likelihood between highest and second peak	F	4	11
A	Light-curve range	L	5	7
$2\sqrt{\beta_1^2 + \beta_2^2}$	Amplitude of periodic component	M	6	2
$\log \theta_2$	Log of hyper-parameter θ_2	M	7	6
$\sigma(\bar{m})$	Standard deviation of residuals about $\bar{m}$	L	8	4
f_1	Best-fit frequency	M	9	1
$\sigma(R_q)$	Standard deviation of residuals from piecewise quadratic fits	M	10	13
Q_1	Best-fit log likelihood	F	11	15
$\overline{m}$	Unweighted mean magnitude	L	12	3
Q_b	Baseline value of frequency spectrum (10th percentile)	F	13	14
$\sigma(R_{\rm model})$	Standard deviation of the best-fit model residuals	L	14	10
N	Number of measurements	L	15	18
θ_1	Hyper-parameter θ_1	M	16	8
$\sigma(\beta_1)$	Posterior uncertainty of β_1	M	17	17
$\sigma(\beta_2)$	Posterior uncertainty of β_2	M	18	16

Note. — [a] Source of parameter = F—frequency spectrum, L—light curve, M—model. [b]Rank in importance for classification as Mira or discrimination between C- and O-rich subtypes. Rank is determined by the mean decrease in the Gini index.

We trained the RF classifier by building 400 decision trees with simulated light curves, each composed of 1.2×10^5 non-Miras and 2.5×10^3 Miras. This ratio was chosen to match the estimated fraction of Mira candidates in the actual data, derived from the ratio of Miras to other stars in the OGLE catalog (after applying the M33 completeness function). We then applied the trained classifier to the actual M33 data and obtained the voted probabilities for each star being a Mira (P_M); Figure 4.7 shows the resulting histogram. Based on a five-fold cross-validation on the simulated light curves, the Mira recovery rate at P_M = 0.5 is 75.4%, and the impurity is 0.7%. There are 5480 objects with probabilities above this value, 5145 of which have A_P > 0.6 mag and were selected for further study.

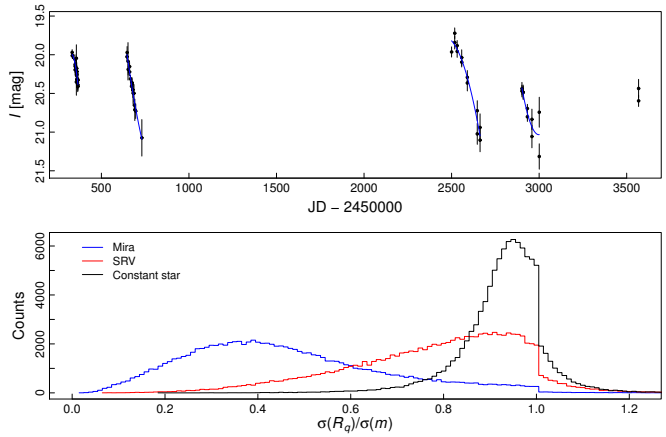

Figure 4.6: Top: piecewise quadratic fit to a simulated Mira light curve. Bottom: distribution of a classification feature based on such fits for simulated Miras (blue), SRVs (red), and constant stars (black).

Figure 4.8 shows three representative light curves for Mira candidates with different values of P_M. The full set of light curves is available in the online edition of Yuan et al. (2017). The Mira subtype was tentatively inferred by using another RF classifier trained on the same features that yielded the probability of each candidate being O-rich (P_O). Using the features of Mira candidates in the LMC bar to classify variables in the inner disk of M33 should be a robust approach, given the similar chemical abundances of both regions (Romaniello et al., 2008; Bresolin, 2011). Figure 4.9 shows the separation between subtypes based on the features with the highest rank in terms of discrimination: P and A_P. The difference in the distribution of variables between the left and middle panels is due to the shallower depth and sparser sampling of the M33 survey relative to the OGLE coverage of the LMC. We caution that this is a limited two-dimensional view of a classification process that is based on 18 features. Figure 4.10 attempts to provide additional insight into the RF classification process by plotting the distribution of a subsample of candidates in other two-dimensional slices of parameter space. Based on a five-fold cross-validation on the simulated light curves, the O-rich recovery rate at P_O = 0.5 is 91.4% and the impurity

is 12.8%, while the corresponding values for C-rich variables are 82.3% and 12.1%.

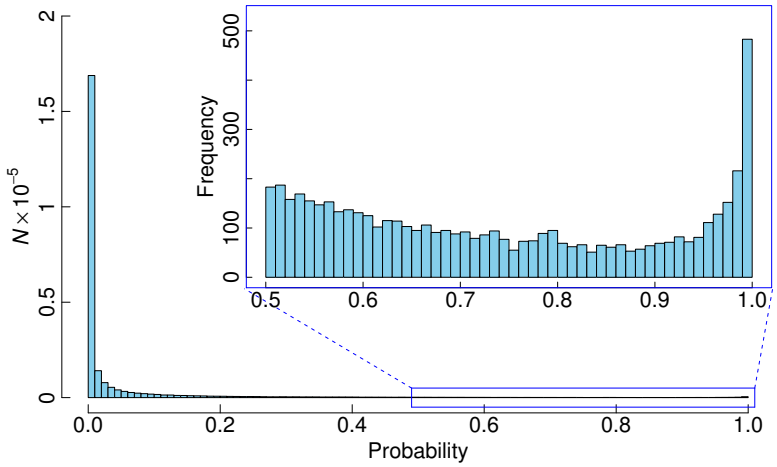

Figure 4.7: Distribution of RF-voted values of Mira probability (P_M) for the entire M33 sample. There are 5480 objects with $P_M > 0.5$.

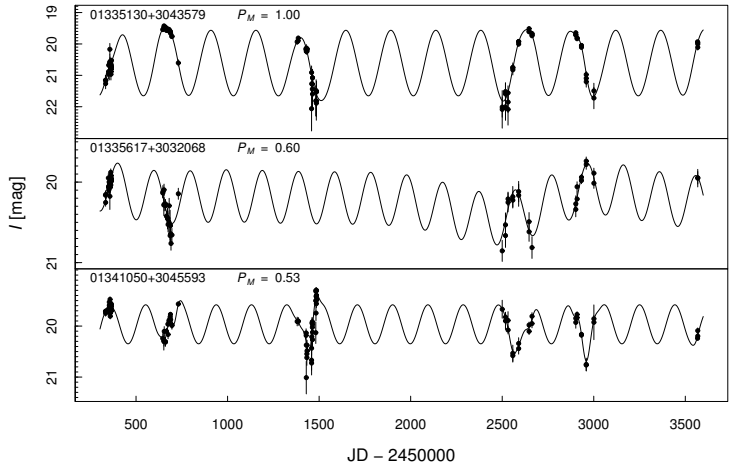

Figure 4.8: Example light curves and best-fit models (solid lines) for likely Miras in M33 with different values of P_M.

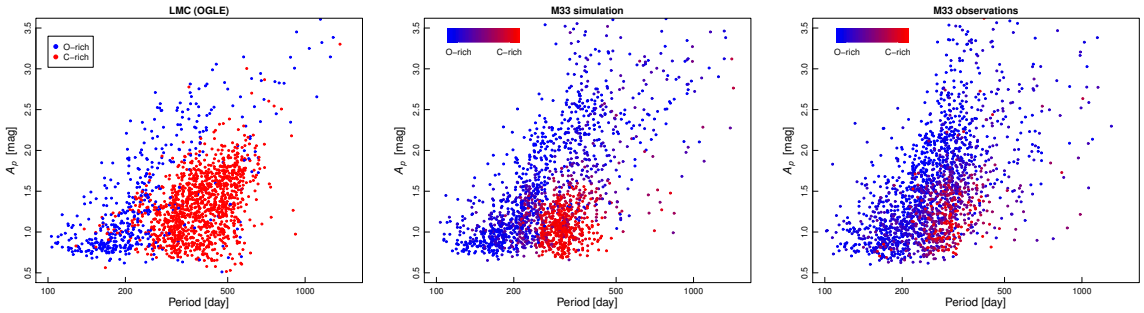

Figure 4.9: RF classification of Mira candidates into O-rich (blue) or C-rich (red), plotted as a function of P and A_P. Left: LMC variables classified by Soszyński et al. (2009). Middle: simulated M33 variables based on the LMC sample but accounting for the shallower depth in absolute magnitude of our survey. Right: Mira candidates in M33 from this work.

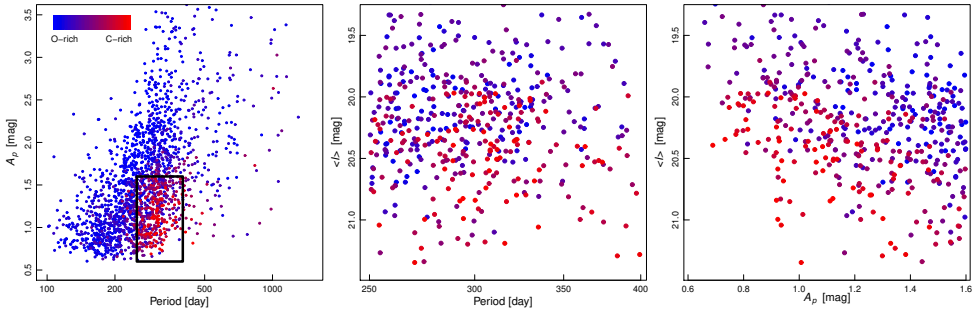

Figure 4.10: Illustration of how multiple attributes help to discriminate O-rich from C-rich Mira candidates. Left: same as right panel of Figure 4.9 but indicating the area of interest where both subtypes overlap. Middle and right: separation of candidates on other two-dimensional slices of the RF parameters.

4.4.2 Comparison with Other Classification Methods

Although we chose RF as our classifier, it is insightful to compare its performance against other popular classifiers. We selected three state-of-the-art classifiers: sparse linear discriminant analysis with l_1 penalty (LDA), sparse logistic regression with l_1 penalty (SLR), and a ν-classification support vector machine with radial basis kernel (SVM). We used the same input features discussed in Section 4.4.1, normalized to zero mean and unit variance.

First, we considered the classification task of Mira versus non-Mira. The comparison was in terms of the receiver operating characteristic (ROC) and its summary statistic, area under the curve (AUC). They were computed via repeated splitting of the simulated data set. In each instance, 10^4 Mira and non-Mira light curves (3.7% of the total) were sampled without replacement to serve as training data, while the rest served as test data. The procedure was repeated 200 times, and the final prediction for each light curve was calculated from the averaged probability across all iterations. The resulting ROC curves, shown in the left panel of Figure 4.11, are nearly identical, with AUC values of 0.984, 0.979, 0.975, and 0.976 for RF, LDA, SLR, and SVM, respectively.

We carried out a similar comparison for the task of classifying Miras into C-rich versus O-rich, with ROC curves plotted in the right panel of Figure 4.11. In this case, RF is significantly superior to the other methods, with AUCs of 0.912, 0.793, 0.787, and 0.801 for RF, LDA, SLR, and SVM, respectively.

4.4.3 Mira Candidates and PLRs

We examined the distribution of the best-fit periods for the selected objects and found a large peak at $P \sim 340$ day that is not seen in the LMC samples. Visual examination of the light curves in this period bin showed that they exhibit a significant change in the mean magnitude of segments obtained from different telescopes. Further examination of

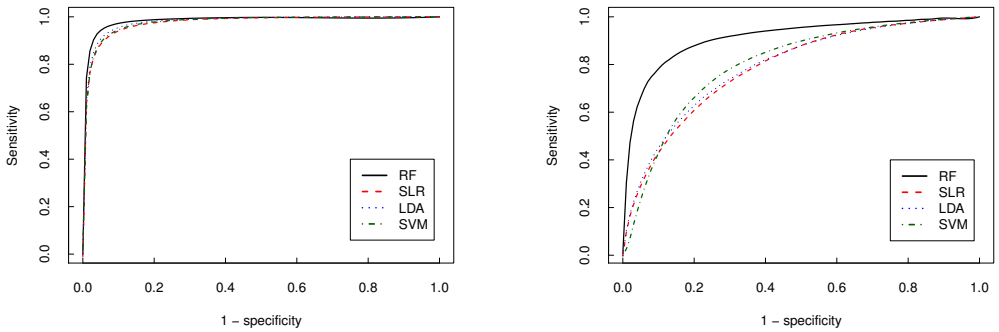

Figure 4.11: ROC curves for classification between Mira/non-Mira (left) and C/O-rich (right) for various classifiers. RF: black solid line; SLR: red dashed line; LDA: blue dotted line; SVM: green dashed-dotted line. While all classifiers have very similar AUCs for the first classification, RF significantly outperforms the others in the latter.

the reference images for each field revealed that, for these objects, the poorer angular resolution of the FLWO images resulted in the blending of several sources (clearly separated in the WIYN frames). We visually inspected each light curve and its respective reference images and rejected affected objects.

Our final sample consists of 1847 Mira candidates. Table 4.2 lists the following properties of the candidates (and their uncertainties, when applicable): IAU-standard ID, coordinates, most likely period, mean I mag, amplitude of the periodic component (A_P), brightest magnitude of the best-fit model light curve (I_m), subtype (O/C), number of light-curve measurements (N), range of magnitudes (A), times (Δt) spanned by the light curve, and time of maximum light for the periodic component (T_0). We classified 1581 and 266 objects as O-rich and C-rich, respectively.

Table 4.2. Mira Candidates in M33

ID (M33SSSJ)	R.A. (J2000)	Decl. (°)	P (day)	$\sigma(P)$ (day)	$\bar{I}$	$\sigma(\bar{I})$	A_P	$\sigma(A_P)$ (mag)	$I_{\min}$	$\sigma(I_{\min})$	O/C	N	A (mag)	$\sigma(A)$ (mag)	Δt (day)	T_0 (day)
01321114+3032588	23.04642	30.54967	324.1	2.0	19.85	0.14	2.12	0.29	18.60	0.36	O	39	3.02	0.25	1955.9	2106.1
01321450+3019349	23.06041	30.32637	309.7	10.0	20.05	0.07	1.70	0.09	19.55	0.08	O	46	2.17	0.15	1914.0	1993.3
01321654+3025260	23.06890	30.42388	295.9	11.3	20.07	0.05	0.78	0.05	19.55	0.05	C	56	1.38	0.11	1955.8	2074.7
01321897+3031226	23.07905	30.52294	258.2	11.0	20.28	0.19	1.88	0.23	19.19	0.20	O	19	2.21	0.33	482.9	2687.5
01322179+3034063	23.09078	30.56841	354.9	6.4	19.48	0.12	3.42	0.22	18.87	0.10	O	16	1.34	0.07	468.8	2613.6
01322351+3030590	23.09797	30.51638	266.0	0.9	20.69	0.10	1.39	0.24	19.80	0.29	O	51	2.29	0.36	1914.0	2069.9
01322586+3033489	23.10773	30.56359	130.0	3.4	20.53	0.04	0.77	0.06	20.05	0.09	O	55	1.21	0.14	1955.9	2078.7
01322828+3017589	23.11782	30.29970	399.8	11.7	18.56	0.20	2.95	0.39	17.65	0.35	O	14	2.04	0.05	1948.8	2167.8
01322948+3026495	23.12283	30.44707	313.4	1.4	20.43	0.08	1.62	0.10	19.68	0.11	O	58	2.24	0.44	1955.8	1955.3
01322979+3034179	23.12411	30.57163	339.6	2.9	19.98	0.07	0.97	0.09	19.44	0.02	C	54	1.66	0.12	1955.9	2072.8
01323105+3031442	23.12939	30.52895	191.0	0.9	20.75	0.09	1.51	0.13	19.74	0.13	O	53	2.50	0.48	1955.9	2006.2
01323349+3038395	23.13955	30.64431	385.8	5.7	18.55	0.04	0.86	0.12	18.15	0.15	O	47	0.91	0.03	1955.9	2166.8
01323465+3032326	23.14437	30.54238	411.6	9.4	20.31	0.10	1.58	0.12	19.74	0.03	C	16	1.29	0.13	468.8	2628.2
01323861+3019003	23.16088	30.31675	153.8	1.1	21.25	0.12	1.82	0.16	20.14	0.18	O	44	2.76	0.92	1955.8	1965.1
01324006+3023433	23.16690	30.39537	373.1	12.9	20.63	0.19	2.38	0.35	19.59	0.26	O	20	2.87	0.51	482.9	2891.0

Note. — Only a portion of this table is shown here to demonstrate its form and content. A machine-readable version of the full table is available at Yuan et al. (2017).

Figure 4.12 shows histograms of periods and amplitudes for both subtypes, while Figure 4.13 shows their deprojected galactocentric distribution. Our survey is limited to the innermost $\sim$5 kpc of the galaxy, and within this limited area we see no statistically significant difference in the distribution of candidates by subtype or period.

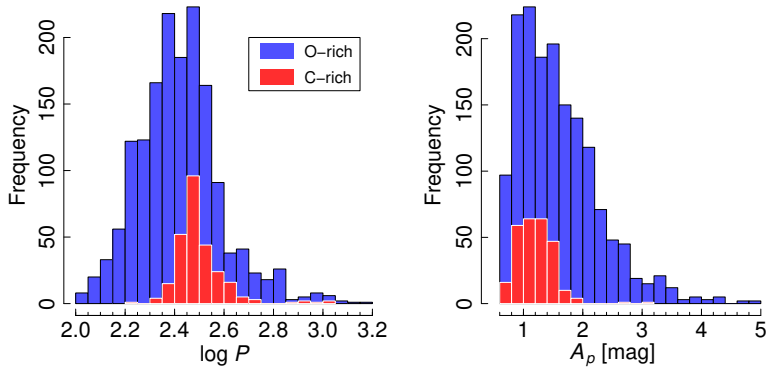

Figure 4.12: Distribution of periods (left) and amplitudes (right) for Mira candidates of each subtype.

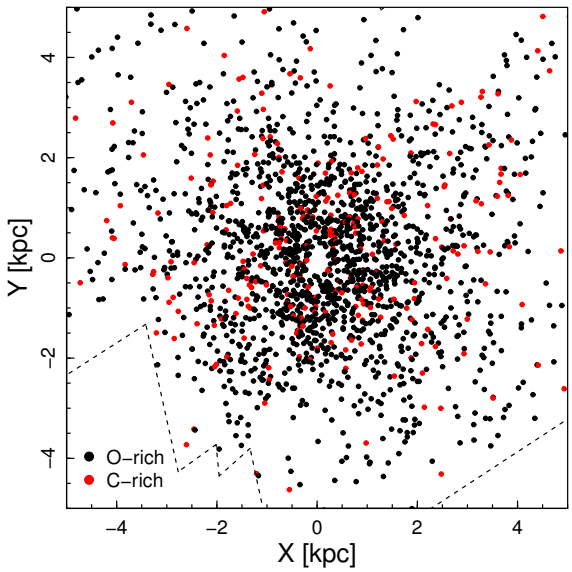

Figure 4.13: Deprojected distribution of Mira candidates (O-rich in black, C-rich in red). The dashed lines indicate the boundaries of our survey.

We note that we recovered the only spectroscopically confirmed Mira in M33 (Barsukova et al., 2011), for which we found $P = 578 \pm 32$ day (in contrast to the previously published estimate of $P = 665$ day). Our classifiers yielded $P_M = 0.89$ and $P_O = 0.9$ for this object.

Figures 4.14 and 4.15 show preliminary PLRs for O-rich Mira candidates in the LMC and M33 at wavelengths ranging from 0.8 to 4.5 μm. We emphasize that the following is a simple analysis to demonstrate the validity of our methods for identifying, phasing, and classifying Mira candidates in M33. A complete analysis (including C-rich candidates) will be presented in a future paper.

The I_m magnitudes of the LMC Mira candidates were determined using the method described in Section 4.2, while the random-phase magnitudes at longer wavelengths were obtained from the SAGE catalog (Meixner et al., 2006). We chose to plot the I_m PLR to show a minimally biased comparison of the relations at this wavelength, since the $V - I$ colors necessary to generate a Wesenheit-corrected mean-light I-band PLR are not available for M33. We show one example of a relation corrected for interstellar extinction for K_s, using the formulation of Soszyński et al. (2009). We note that this formulation may not be appropriate to correct for the circumstellar dust that is especially prevalent among C-rich and long-period Miras (see Ita & Matsunaga (2011) for a thorough analysis of this issue). We solved for quadratic PLRs,

$$m = a_0 + a_1(\log P - 2.3) + a_2(\log P - 2.3)^2 \,, \tag{4.3}$$

using an iterative 3σ clipping and removing the single largest outlier in each band until convergence.

The M33 sample was restricted to 1161 candidate variables with $A_P/A < 1.1$, $\sigma(A_P)/A_P < 0.15$, $\sigma(P)/P < 0.1$, and $P < \Delta t$. These selection criteria were based on

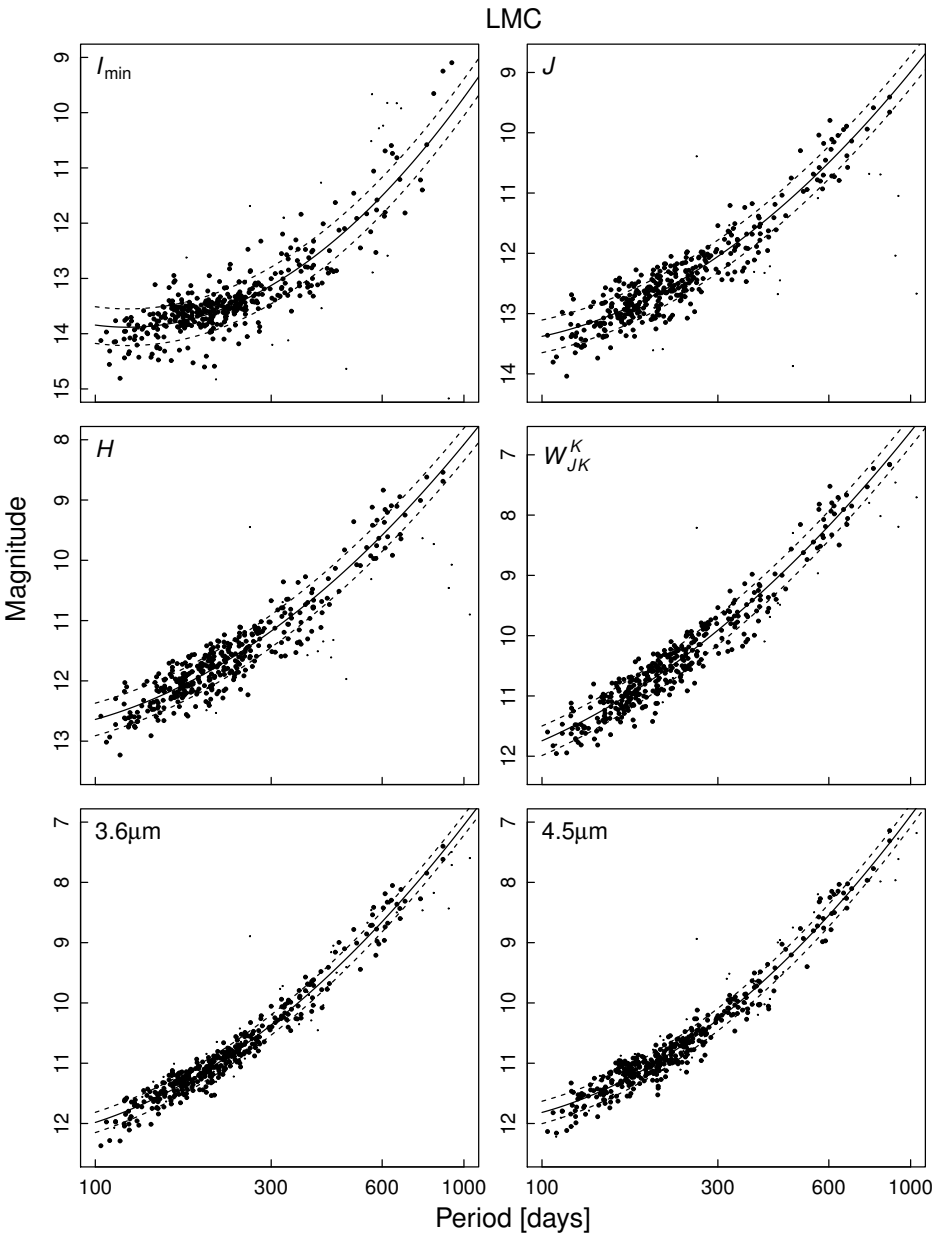

Figure 4.14: PLRs in several bands for Mira candidates classified as O-rich in the LMC. The solid lines show the best-fit quadratic relations to the final LMC samples (large symbols) after iterative 3σ clipping of outliers (small dots). Dashed lines indicate the 1σ dispersion in the fits.

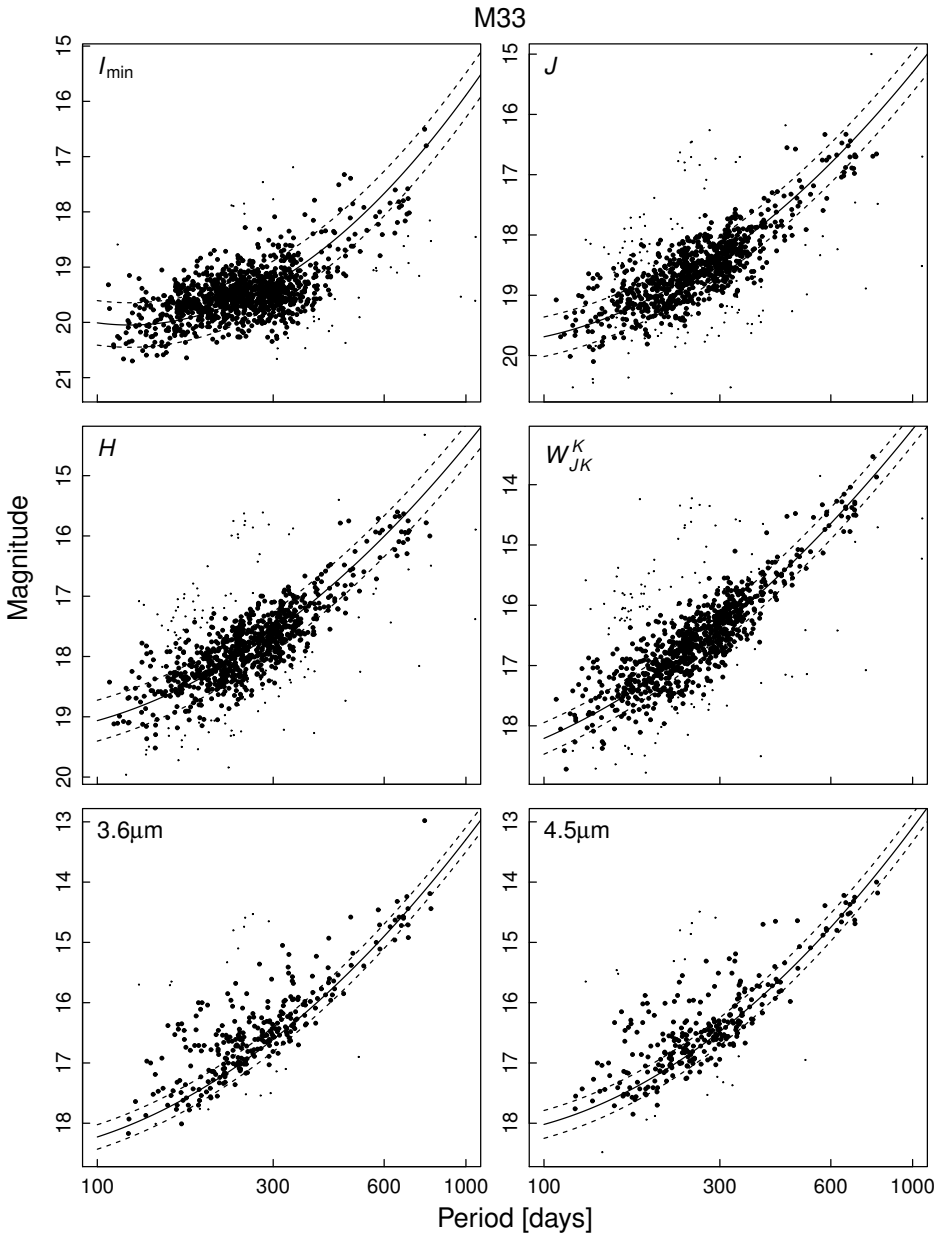

Figure 4.15: PLRs in several bands for Mira candidates classified as O-rich in M33. The solid lines show the LMC-based quadratic relations of Figure 4.14 shifted by the best-fit relative distance modulus in each band (including blending correction). Small dots indicate variables removed by iterative 3σ clipping. Dashed lines indicate the 1σ dispersion of the Gaussian component of the model.

an examination of the input and recovered parameters for the simulated M33 Miras. When the amplitude of the periodic component significantly exceeds the range of magnitudes spanned by the data and/or the best-fit period is longer than the time span of the light curve, the recovered parameters exhibit considerably larger scatter and the fraction of variables with successfully recovered periods (as defined in H16) is noticeably lower. The simulated O-rich Miras that met our selection criteria had input/output ratios of A_P of 1.02 ± 0.21 versus 0.70 ± 0.82 for the others. Likewise, the fraction of successfully recovered periods was 86% for the variables meeting the criteria versus only 45% for the others.

The I_m magnitudes of the M33 Mira candidates are from Table 4.2, while the random-phase magnitudes at longer wavelengths were taken from Javadi et al. (2015, for JHK_s) and Thompson et al. (2009, for 3.6 and 4.5 μm). We matched the catalogs using tolerances of 0.3″and 0.5″and found 972 and 302 counterparts, respectively. We fixed the linear and quadratic terms of the PLRs to those derived from the LMC sample and solved for a_0, applying an iterative 3σ clipping that removed the single largest outlier at a time. Once this procedure converged, we modeled the cumulative distribution of the PLR residuals as the combination of a Gaussian (to account for the finite width of the instability strip) plus an exponential distribution toward brighter values (to account for blends, which can only bias the residuals in one direction). The final values of a_0 include these "blending corrections," which amount to 0.09 and 0.25 mag at near- and mid-infrared wavelengths, respectively. The larger contamination for the two longest bands is likely due to the significantly poorer angular resolution of the *Spitzer* images. The scatter in the M33 PLRs (after accounting for blended objects) compares favorably with the higher-quality LMC samples, and the mean (error-weighted) LMC-relative distance modulus of 6.31 ± 0.11 mag is consistent with previous determinations (Bonanos et al., 2006, PM11).

4.5 Summary

We carried out a search for Mira variables in M33 using sparsely sampled I-band light curves. We determined periods using a novel semi-parametric Gaussian process model and used the RF method to identify Mira candidates and classify them into C- or O-rich subtypes. We identified 1847 likely Mira candidates, most of them O-rich, that exhibit PLRs with dispersions comparable to those seen in the LMC.

5. NIR PROPERTIES OF MIRA CANDIDATES IN THE LMC

In this section, we study the near-infrared properties of 690 Mira candidates in the central region of the Large Magellanic Cloud (LMC) based on time-series observations in JHK_s. We compared the observed colors for Oxygen-rich (O-rich) and Carbon-rich (C-rich) Mira subtypes, and hypothesized that these two subtypes have similar intrinsic $J-H$ and $H-K_s$ color indices, while the light from C-rich objects are strongly attenuated by circumstellar dust, with a different reddening law from interstellar dust. We derived empirical Period-Luminosity Relations (PLRs) for the O-rich subtype, with the K_s-band scatter as low as 0.12 mag. We refer to confirmed Miras and unconfirmed Mira candidates collectively as Miras throughout this section.

5.1 Data

This study made use of I-band measurements of Miras from OGLE-III (Soszyński et al., 2009) and JHK_s measurements from the LMC Near-Infrared Synoptic Survey (LMCNISS; Macri et al., 2015).

5.1.1 LMCNISS Measurements

LMCNISS observed the central region of the LMC in JHK_s using the CPAPIR camera at the CTIO 1.5m telescope. The detailed description of the survey and data reduction can be found in Macri et al. (2015). Based on the locations of the LMC Miras, we extracted 681, 679, and 676 light curves in JHK_s, respectively. There are 668 objects with a total number of 84,852 measurements in all 3 bands. The measurements were separated into three groups based on their observation dates, with group 1 observed in 2006 November, group 2 observed in 2007 January and group 3 observed in 2007 November. Compared to the periods of Miras, the time spans for any of the groups are short enough to reject

extreme outliers based on the mean magnitude within the group. Across this study, "extreme outliers" are defined as data beyond outer fences, which are mathematically equal to $Q_1 - 3 \times (Q_3 - Q_1)$ and $Q_3 + 3 \times (Q_3 - Q_1)$ with Q_1 and Q_3 being the values at the 25 and 75 percentiles, respectively. With extreme outliers rejected, the median number of measurements is 42, 44, and 38 per Mira in JHK_s, respectively.

5.1.2 OGLE-III Measurements

We retrieved the I-band Mira light curves from the OGLE-III database for those variables observed by LMCNISS. The OGLE-III measurements are characterized by dense sampling and a low noise level. The observation baseline is as long as 12 years for most Miras, and gracefully covers the NIR observation dates from LMCNISS. We visually inspected the light curves and rejected 46 out of 551,969 measurements. The rejected measurements noticeably deviate from close-by measurements and are possibly due to occasional photometry artifacts.

5.2 The LMC Mira Color Excess

The O-rich and C-rich subtypes of Miras are known to exhibit different NIR colors (Soszyński et al., 2009), which is possibly due to the different circumstellar environments (Ita & Matsunaga, 2011). We studied the circumstellar extinction of the LMC Miras using the $J - H$ and $H - K_s$ color indices. We obtained mean color indices for each Mira by firstly calculating them in each temporal group separately, then averaging them over the three groups with equal weights. Figure 5.1 gives an example of the relation between mean $J - H$ color index and JH measurements. Our O-rich/C-rich classification is based on Soszyński et al. (2009), which separated the two subtypes in the $W_I - W_{JK}$ diagram.

We found that the LMC Miras clustered in two regions in the observed color-color diagram as shown in Figure 5.2. Most O-rich Miras (blue circles) are centered at [$0.345_{\pm 0.005}$,

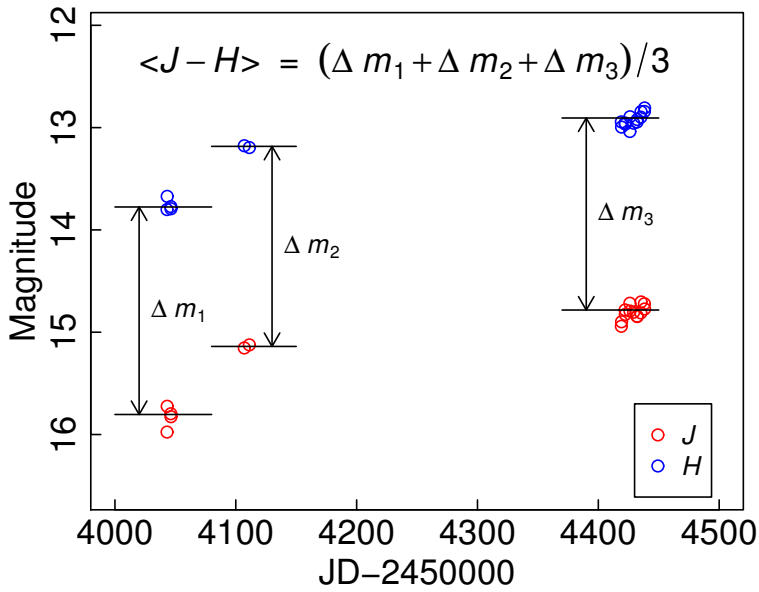

Figure 5.1: Example of how the mean $J - H$ color index is derived from the three groups of measurements for Mira candidate OGLE-LMC-LPV-08476.

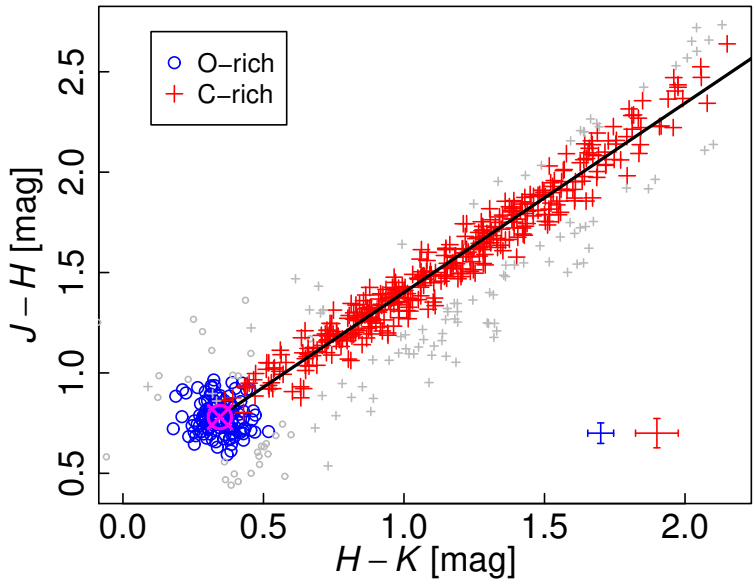

Figure 5.2: Observed color-color diagram for LMC Miras. The O-rich and C-rich subtypes are indicated by blue circles and red pluses, respectively. Outliers are indicated by grey symbols. The magenta circled cross shows the centroid of O-rich Mira color indices. The black line is the first-order best-fit for the C-rich Mira colors.

$0.780_{\pm 0.006}$], while the C-rich Miras (red pluses) are located along a narrow strip. We obtained the centroid of the O-rich Mira cluster, excluding points outside of a 0.2 mag radius. We fit a straight line to the C-rich Miras iteratively, with the first iteration excluding any points further than 0.2 mag from the best-fit line, and the second iteration excluding extreme outliers. Interestingly, the best-fit line

$$(J-H)-0.780 = 0.943_{\pm 0.010} \times [(H-K)-0.345] + 0.003_{\pm 0.009}$$

passes through the centroid of O-rich Mira cluster well within its uncertainty. This suggests the C-rich Miras might have similar intrinsic color indices to the O-rich ones.

To confirm the intrinsic colors of C-rich Miras, we obtained model-based color indices for C-rich giants. Aringer et al. (2009) computed synthetic photometry for C-rich giants using hydrostatic models and provided intrinsic luminosities in various filters, including 2MASS JHK_s. We restricted their results to those with $\log(g) = 0$, $Z = Z_{\odot}$, and $M = 1M_{\odot}$ or $2M_{\odot}$. The model colors are shown in Figure 5.3. We found that the model colors for C-rich giants are consistent with the observed colors of O-rich Miras, which supports the assumption that two subtypes of Miras have similar intrinsic color indices, while only the C-rich Miras are surrounded by a considerable amount of circumstellar dust that noticeably reddens the light. We also noticed that the observed color indices for C-rich Miras are correlated with their pulsating periods, as shown in Figure 5.4.

We compared our color measurements to those of Galactic C-rich variable stars (Whitelock et al., 2006), as well as the extinction laws of interstellar dust. With the assumption that O-rich and C-rich Miras have the same $J-H$ and $H-K_s$ color indices, the color excess ratio $E(J-H)/E(H-K_s)$ for C-rich Miras simply equals the slope in the color-color diagram, which is 0.94 ± 0.01. We transformed the photometric measurements of Whitelock et al. (2006) to 2MASS JHK_s magnitudes with Equations 33–36 of Carpen-

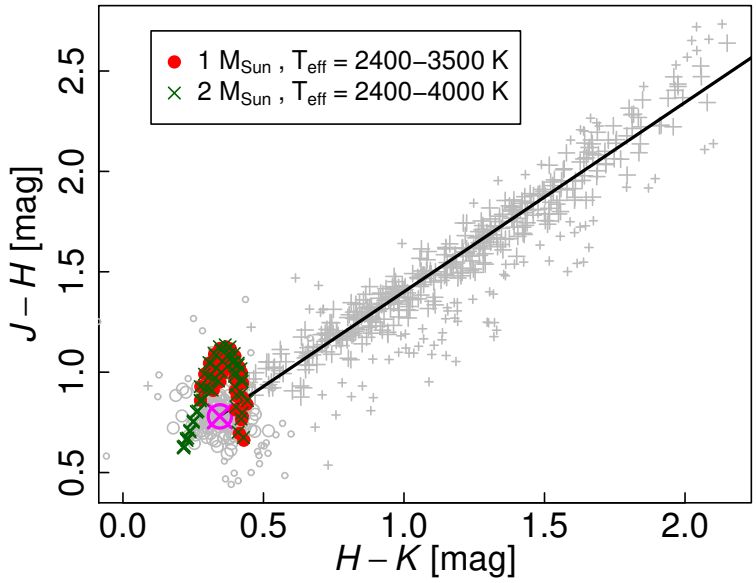

Figure 5.3: Same as Figure 5.2 but showing the intrinsic color indices of C-rich giants from hydrostatic models (Aringer et al., 2009). Red filled circles show $M = 1M_{\odot}$ models while the green crosses indicate $M = 2M_{\odot}$ models. The observed color indices of C-rich and O-rich Miras are shown in gray pluses and circles, respectively.

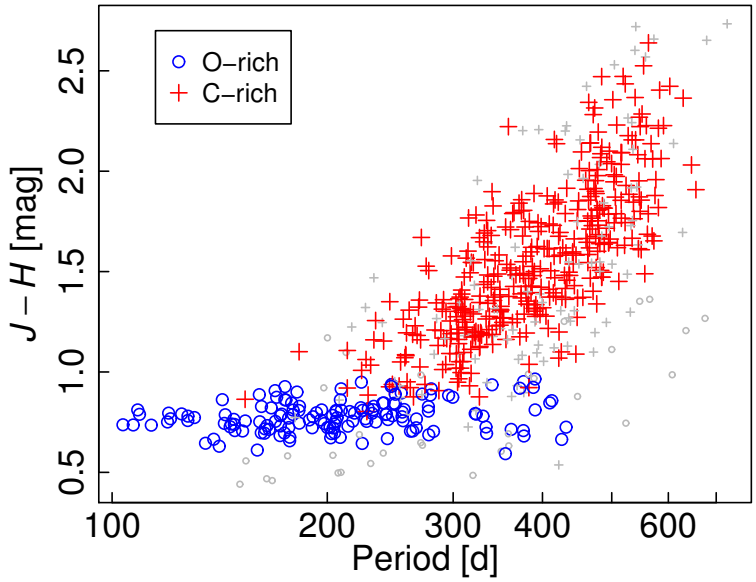

Figure 5.4: Correlation between $J - H$ color index and period for C-rich Miras (red pluses). The O-rich Miras are indicated by blue circles for comparison.

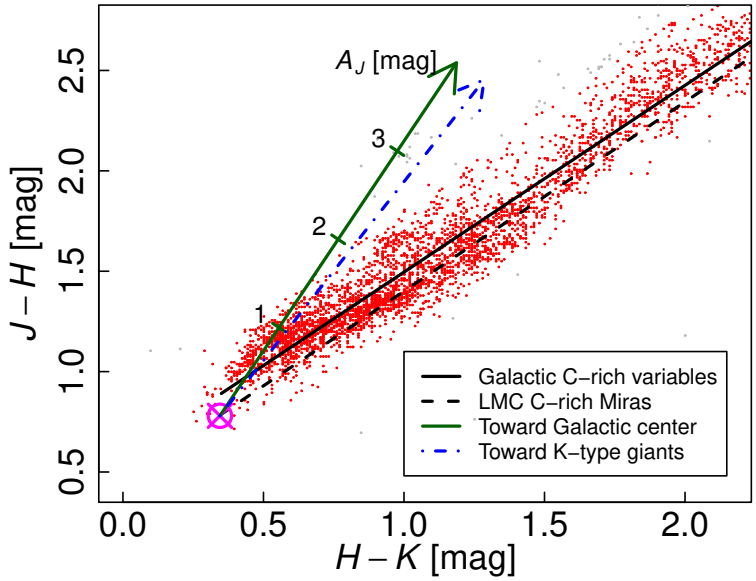

Figure 5.5: Observed color indices for Galactic C-rich variable stars (red dots) from Whitelock et al. (2006), direction of reddening toward Galactic center (green solid arrow, Nishiyama et al., 2009), and direction of reddening toward Galactic K-type giants (blue dashed arrow, Wang & Jiang, 2014). The best first order best-fit of red dots is shown in black solid line. The black dashed line indicates the best-fit for the LMC C-rich Miras.

ter (2001) and obtained a similar color excess ratio ($0.930_{\pm 0.003}$), as shown in Figure 5.5. This value, however, is very different from the interstellar extinction laws. For example, Nishiyama et al. (2009) found $E(J-H)/E(H-K_s) = 2.09$ towards the direction of Galactic center; Wang & Jiang (2014) derived a value of 1.78 using a sample of 5942 K-type giants. This implies different grain sizes between the circumstellar dust around C-rich stars and interstellar dust of various environments.

5.3 Template Mira Light Curves and PLRs in the NIR

While the I-band light curves of the LMC Miras are densely sampled, the NIR observations only cover three "epochs" in terms of sampling long-period variables. To obtain the mean JHK_s magnitudes, we derived template light curves at these wavelengths from

the I-band light curves, as described below.

5.3.1 Templates Light Curves

We constructed least-squares regression models to calculate the $I-J$, $I-H$, and $I-K_s$ curves from a set of predictors that can be estimated at any epoch within the I-band coverage. The predictors include time-dependent quantities such as phase ϕ and absolute I-band magnitude $I_{\rm abs}$, as well as time-independent quantities such as period P, mean color indices, and I-band amplitude of the periodic component A_I. The JHK_s curves were then obtained by subtracting $I-J$, $I-H$, and $I-K_s$ values from the I-band light curves, respectively. The periods were derived by applying the periodogram technique described in He et al. (2016) to the I-band measurements, and the phases are relative to the time of maximum light in I-band.

We derived the NIR templates for JHK_s bands and O-rich/C-rich subtypes separately, and here we take the model of J-band O-rich Miras as an example to illustrate the procedure. We began by obtaining I-band model curves by fitting the semi-parameter Gaussian process model (He et al., 2016) to the I-band measurements. Any sampling gaps in the I-band observations longer than 50 days were excluded from the model curves, and as a result the selection of fitting methods does not have a noticeable impact on the shape of model I curves. We then calculated the predictors and trained a regression model with the entire $I-J$ measurements of all the O-rich Mira light curves simultaneously. To account for possible nonlinearity in the aforementioned time-dependent or time-independent quantities, we increased the number of predictors by adding higher-order and nonlinear functions of these quantities, which are listed in Table 5.1. We selected a subsample of these predictors using the LASSO algorithm (Tibshirani, 1994) in order to avoid overfitting. The regularization factor was determined from a 10-fold cross-validation. With the subsample of predictors, we trained the least-squares regression for two iterations, with

the second iteration excluding extreme outliers of residuals from the first iteration. The best-fit coefficients for all the six combinations are presented in Table 5.1. The typical scatter of residuals from best-fit models is 0.12 mag.

Table 5.1: Regression Predictors and Coefficients

Predictor	O-rich coefficients			C-rich coefficients		
$(I-)$	J	H	K_s	J	H	K_s
$*I_{\rm abs}$	7.9e-01	7.4e-01	7.3e-01	3.4e-01	3.7e-01	4.4e-01
$*I_{\rm abs}^2$	NA	NA	NA	-2.0e-03	-1.1e-02	-3.5e-03
$*F = 10^{-0.4 I_{\rm abs}}$	2.1e-03	NA	NA	NA	NA	-4.8e-04
$*F^2$	NA	1.4e-05	8.9e-06	2.8e-06	NA	NA
$*F^3$	-1.2e-08	-3.7e-08	-2.0e-08	NA	NA	NA
$*F^{0.5}$	NA	NA	NA	NA	NA	-4.5e-02
$*\cos(2\pi\phi)$	NA	-8.4e-03	-4.0e-02	6.1e-02	4.6e-02	4.3e-02
$*\sin(2\pi\phi)$	1.9e-01	2.0e-01	2.1e-01	7.2e-02	8.3e-02	7.7e-02
$*\cos(4\pi\phi)$	1.4e-02	1.1e-02	1.5e-02	-3.5e-03	-1.9e-03	NA
$*\sin(4\pi\phi)$	9.4e-03	1.6e-02	1.8e-02	1.7e-02	1.8e-02	1.8e-02
$*\cos(6\pi\phi)$	1.7e-02	7.7e-03	7.7e-03	5.2e-03	6.6e-03	NA
$*\sin(6\pi\phi)$	NA	NA	NA	-4.8e-03	-9.0e-03	-1.3e-02
P	-7.7e-02	1.9e-02	3.4e-02	2.5e-03	5.4e-03	5.4e-03
$\log P$	6.3e+03	-3.2e+03	-5.5e+03	NA	-2.6e+01	-3.3e+01
$\log^2 P$	-1.1e+03	4.8e+02	8.4e+02	NA	NA	NA
$\log^3 P$	1.1e+02	-4.5e+01	-7.9e+01	NA	NA	NA
$\log^{0.5} P$	-9.5e+03	5.0e+03	8.7e+03	NA	7.7e+01	1.0e+02
$J - H$	-5.6e-01	3.9e-01	3.5e-01	-3.5e-01	4.8e-01	3.2e-01
$H - K$	-5.4e-01	-4.8e-01	4.1e-01	NA	-2.4e-01	3.9e-01
A_I	7.9e-03	1.7e-02	1.7e-02	5.0e-02	7.4e-02	1.0e-01
Intercept	4.2e+03	-2.3e+03	-4.0e+03	2.7e+00	-5.4e+01	-7.3e+01
σ [mag]	0.12	0.12	0.12	0.12	0.13	0.15

Note. — Predictors beginning with $*$ indicates time-dependent quantities.

NA indicates predictors rejected by the LASSO regularization.

With the best-fit regression models, we predicted the color curves and hence the JHK_s curves. For each Mira in each band, we solved an overall offset between the predicted curve and actual measurements. The final NIR template light curves were obtained by subtracting the offsets from the predicted curves. Figure 5.6 shows an example of I-band model curve and NIR template light curves for one Mira. We obtained the mean, maximum, and minimum JHK_s magnitudes from the piece-wise template light curves and list them in Table 5.2.

In the above analysis, we excluded six Miras due to their unusual color indices, and one Mira due to its large residuals. We excluded three Miras from the regression results because they show spikes at some phase in the template light curves, possibly due to incorrect phase determination. There are 22 more Miras that do not have $J-H$ or $H-K_s$ information, and they cannot be used in the regression models. For these 32 Miras without least-squares regression results, we computed their mean magnitudes by taking the average of the mean magnitudes of the individual temporal groups. Then mean magnitudes for these Miras are also listed in Table 5.2 and used in the following analysis.

5.3.2 Mira PLRs in NIR

Miras are not strictly periodic but exhibit chaotic variations in their light curves. Given the limited time coverage of the template light curves for the LMC Miras, it is not possible to obtain their true mean magnitudes. We used the median value of the maximum and minimum magnitude of each piecewise template light curve as the "mean magnitude". For those without template light curves, we used the averaged magnitudes of the three temporal groups as the "mean magnitude".

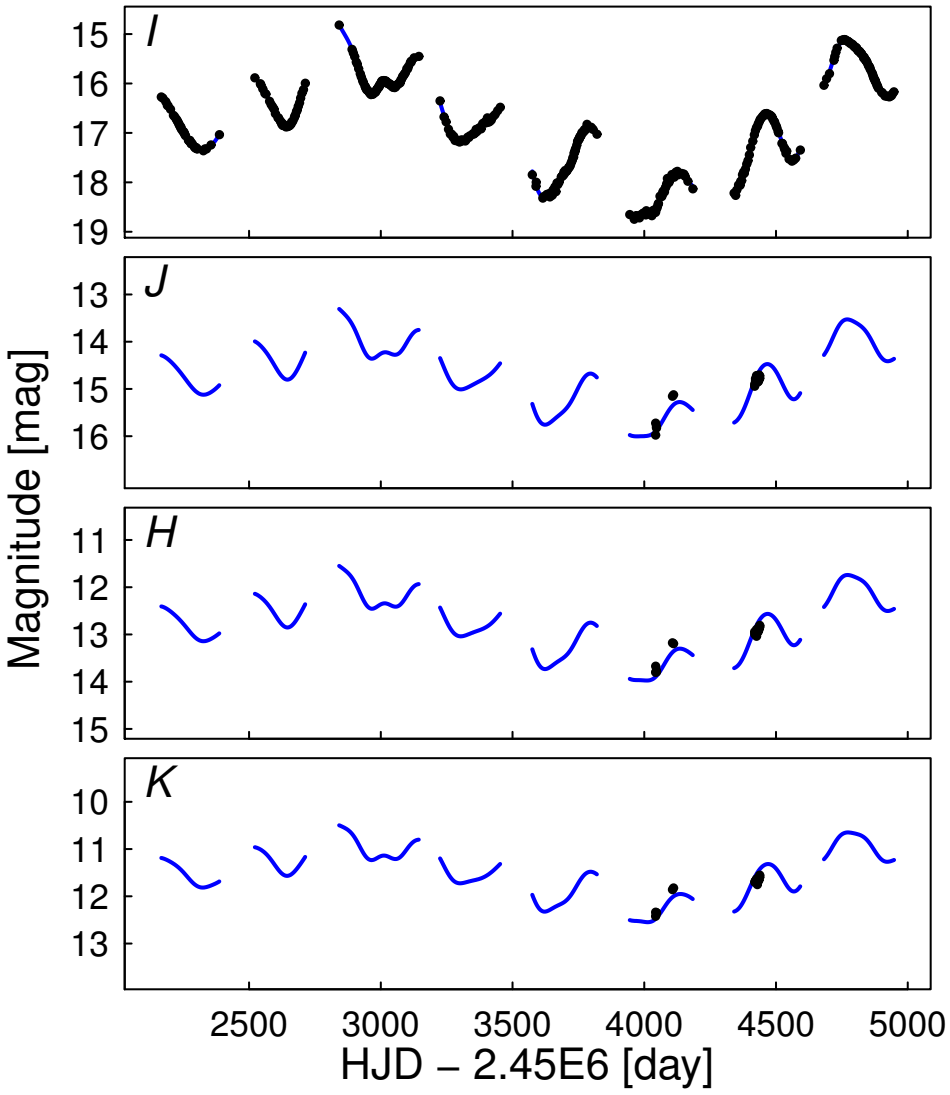

Figure 5.6: Template light curves in $IJHK_s$ (blue curves) for Mira OGLE-LMC-LPV-08476. The measurements are indicated by black points.

Table 5.2: JHK_s Magnitudes for LMC Miras

ID	flag[a]	St[b]	J (mag)			σ	H (mag)			σ	K (mag)			σ	Period	$t^c_{\rm ref}$
			mean	max	min		mean	max	min		mean	max	min		(days)	
8476	0	C	14.72	16.00	13.31	0.11	12.79	13.97	11.55	0.12	11.52	12.55	10.50	0.12	324.4	3466.7
8922	0	C	14.17	14.74	13.42	0.03	12.52	13.03	11.87	0.03	11.26	11.69	10.71	0.05	375.2	3566.9
9071	0	C	13.10	13.68	12.26	0.01	11.52	12.02	10.80	0.03	10.37	10.77	9.81	0.06	463.0	3681.0
9489	0	C	13.89	15.56	12.82	0.05	12.07	13.60	11.13	0.06	10.73	12.04	9.96	0.10	485.9	3419.4
9601	0	C	12.73	13.35	12.18	0.02	11.52	12.03	11.06	0.06	10.72	11.11	10.39	0.10	340.7	3438.9
9653	0	O	12.07	12.52	11.50	0.03	11.36	11.84	10.75	0.08	11.00	11.44	10.45	0.12	333.4	3441.7
9702	0	C	12.47	13.08	12.12	0.04	11.47	11.97	11.17	0.09	10.94	11.31	10.72	0.10	305.5	3511.3
10225	0	C	15.86	17.06	14.74	0.02	13.79	14.93	12.75	0.02	12.03	13.01	11.13	0.02	543.8	3525.2
10601	0	O	12.25	12.51	12.02	0.06	11.36	11.63	11.10	0.06	10.98	11.26	10.73	0.07	278.8	3592.0
10609	0	O	12.94	13.27	12.59	0.03	12.22	12.57	11.83	0.04	11.82	12.14	11.46	0.04	176.7	2784.4
10643	0	O	12.77	13.10	12.49	0.07	12.19	12.54	11.89	0.07	11.79	12.14	11.50	0.08	159.6	3633.7
10702	0	C	12.42	13.58	11.38	0.05	10.78	11.79	9.91	0.06	9.63	10.47	9.00	0.13	537.4	2640.7
10836	0	C	16.91	17.93	15.78	0.10	14.46	15.48	13.36	0.04	12.43	13.29	11.48	0.02	518.7	3725.9
11035	0	O	11.56	12.02	11.10	0.09	10.62	11.11	10.10	0.19	10.34	10.77	9.87	0.15	390.5	2852.0
11105	0	O	12.70	13.03	12.39	0.11	11.77	12.12	11.44	0.04	11.30	11.65	10.99	0.01	223.2	2733.7
11157	0	C	12.59	13.14	12.03	0.04	11.25	11.72	10.79	0.09	10.42	10.76	10.08	0.12	414.2	3004.1
11163	0	C	12.72	13.27	12.27	0.01	11.51	11.96	11.12	0.02	10.84	11.18	10.54	0.04	312.9	2763.2
11237	0	C	14.36	16.32	12.89	0.07	12.56	14.41	11.24	0.16	11.00	12.58	9.89	0.18	515.2	3092.8
11318	0	O	12.79	13.18	12.51	0.01	11.99	12.40	11.69	0.01	11.69	12.10	11.40	0.01	161.8	2847.7
11363	0	O	12.34	12.76	11.90	0.04	11.59	12.03	11.12	0.03	11.18	11.59	10.75	0.04	281.6	2763.2
11557	0	C	13.84	16.50	12.49	0.08	12.23	14.71	11.05	0.06	10.98	13.11	10.04	0.11	423.4	3355.4
11854	0	C	16.03	16.87	15.31	0.06	13.74	14.55	13.06	0.05	11.92	12.62	11.33	0.02	551.3	3042.1
12125	0	O	11.36	11.88	10.82	0.02	10.39	10.94	9.79	0.07	10.17	10.68	9.63	0.04	389.3	2759.0
12236	0	O	13.20	13.55	12.86	0.01	12.49	12.86	12.11	0.02	12.14	12.49	11.79	0.01	146.3	2906.9

Note. — Only a portion of this table is shown here to demonstrate its form and content. The full content are available to interested readers by request.

[a]Flag '0' indicates magnitude from regression model. All the other labels indicate Miras with mean magnitudes calculated in a simpler method. Flag 'c' indicates outliers in color-color diagram; 'r' indicates regression outlier; 'm' indicates $J-H$ or $H-K$ measurements are missing; 'p' indicates Miras with possible improper phases.

[b] Mira subtypes.

[c] HJD - 2450000.

We fit empirical quadratic relations

$$m = a_0 + a_1(\log P - 2.3) + a_2(\log P - 2.3)^2 \tag{5.1}$$

to the periods and magnitudes of O-rich Miras, restricting the periods between 120 days and 1000 days. We rejected extreme outliers in each band during the fit. The O-rich PLRs derived in this work are broadly consistent with the preliminary relations derived in Section 4 based on single-epoch 2MASS observations. However, our relations exhibit $\sim 2\times$ lower dispersion, as shown in Figure 5.7. We summarized the best-fit PLRs in Table 5.3.

We also tested the PLRs using Wesenheit indices based on the interstellar extinction law from Fouqué et al. (2007), though this may be different from the circumstellar dust extinction law for Miras. These Wesenheit indices are

$$\begin{aligned} W_{JH} &= H - 1.611 \cdot (J - H) \\ W_{JK} &= K_s - 0.679 \cdot (J - K_s) \\ W_{HK} &= K_s - 1.900 \cdot (H - K_s) \end{aligned} \tag{5.2}$$

We found that the K_s-band PLR gives the least scatter in all six bands or Wesenheit indices. We caution the readers that the these Wesenheit indices may be inappropriate as distance indicators since the circumstellar dust extinction law and color dependences of PLRs are not fully understood for Miras at present.

We explored the correlations of the PLR residuals across different bands, which is defined as $m_{\rm obs} - m_{\rm PLR}$. For the C-rich Miras, we used the PLRs of O-rich Miras as fiducial C-rich Mira PLRs to compute "residuals". The same method has been used by Ita & Matsunaga (2011) to investigate the circumstellar extinction of Miras. We found strong correlations of the residuals across JHK_s bands for both O-rich Miras and C-rich

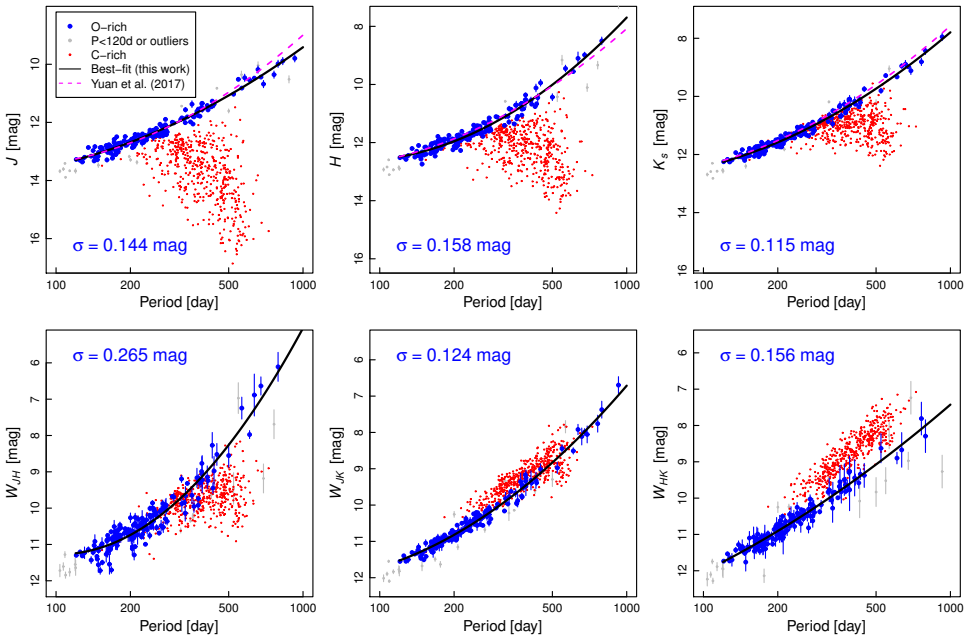

Figure 5.7: PLRs of the LMC Miras. The blue points indicate O-rich Miras while the red points indicates C-rich Miras. The black solid lines show the best-fit quadratic relations to the O-rich Miras, while the magenta dashed lines show the PLRs based on single-epoch 2MASS observations (Yuan et al., 2017).

Table 5.3. O-Rich Mira PLRs in the LMC

Band	a_0	a_1	a_2	σ (mag)	N_i	N_f
J	12.686 ± 0.013	-3.21 ± 0.11	-2.10 ± 0.25	0.144	180	170
H	11.959 ± 0.014	-3.29 ± 0.12	-4.00 ± 0.36	0.158	176	169
K_s	11.585 ± 0.010	-3.61 ± 0.09	-2.58 ± 0.21	0.115	176	169
W_{JH}	10.742 ± 0.024	-3.67 ± 0.20	-6.39 ± 0.61	0.265	174	164
W_{JK}	10.823 ± 0.011	-3.83 ± 0.09	-2.91 ± 0.23	0.124	174	162
W_{HK}	10.910 ± 0.014	-4.11 ± 0.12	-1.24 ± 0.35	0.156	175	164

Note. — Here N_i is the initial number of the variables in the sample (excluding those with periods shorter than 120 days), N_f is the final number after rejecting the extreme outliers, and σ is the Gaussian width.

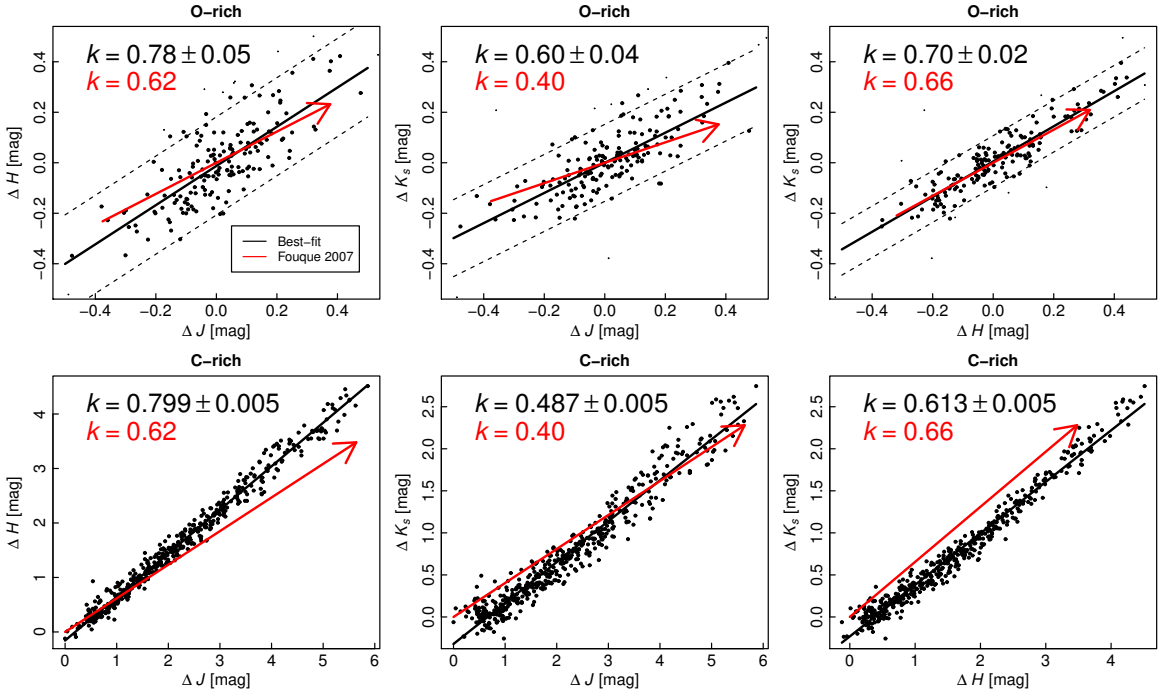

Figure 5.8: Correlations of PLR residuals between different bands for O-rich Miras (upper panels) and C-rich Miras (lower panels) in the LMC. We adopted the PLRs of O-rich Miras to compute the "residuals" for C-rich Miras. The black solid lines indicate the best-fit of correlations of residuals, while the black dashed lines indicate the $\pm 2\sigma$ widths of the relations. The red arrows indicate the direction of interstellar reddening based on Fouqué et al. (2007).

Miras, as shown in Figure 5.8. Several factors may contribute to these correlations: (1) Temperature variations that cause color dependences of PLRs (see the related discussion in Madore & Freedman, 1991); (2) interstellar dust extinction; (3) circumstellar dust extinction; and (4) possibly correlated measurement noise. For the C-rich Miras the factors (2) and (4) are marginal compared to the large variation in residuals, but the justification of using PLRs of O-rich Miras to compute "residuals" remains a question. We compared the slopes of these correlations to the theoretical interstellar reddening vectors derived from Fouqué et al. (2007) (red arrows in Figure 5.8) and found they are statistically different in most cases.

While we cannot break the degeneracy of these many factors, we investigated observed Period-Luminosity-Color Relations (oPLCRs) for C-rich Miras given their tight correlation of residuals. As we have mentioned, in terms of studying the circumstellar dust extinction and color dependences of PLRs for C-rich Miras, the interstellar extinction and correlated measurement errors are negligible. We fit the oPLCRs in the form of

$$M_{JK}^{K} \equiv K_s - b \cdot (J - K_s) = a_0 + a_1(\log P - 2.3) \tag{5.3}$$

where a_0, a_1 and b are free parameters. We applied a revised version of least-squares fit to minimize $\sigma \cdot [(1+b)^2\sigma_K^2 + b^2\sigma_J^2]^{-1/2}$ instead of σ to account for the dependence of σ on b, where σ is scatter of the best-fit model. We used equal values of σ_J and σ_K since the intrinsic scatters are not available. We found that the choice of the values of σ_J and σ_K or the number of orders in the period terms does not significantly change the results. For example, using the scatter ratio of PLRs of O-rich Miras ($\sigma_J/\sigma_K = 1.25$) or adding a quadratic term in the period only changes the value of b by 2.6% and 1.1%, respectively. We excluded any measurements with uncertainties greater than 0.3 mag or extreme outliers during the fit. We also obtained the oPLCRs for other combinations of colors, and list the results in Table 5.4. We show the best-fit oPLCRs in Figure 5.9. We are aware that the circumstellar dust extinction law may be degenerate with intrinsic color terms, and one should not take the M values as reddening-free Wesenheit indices.

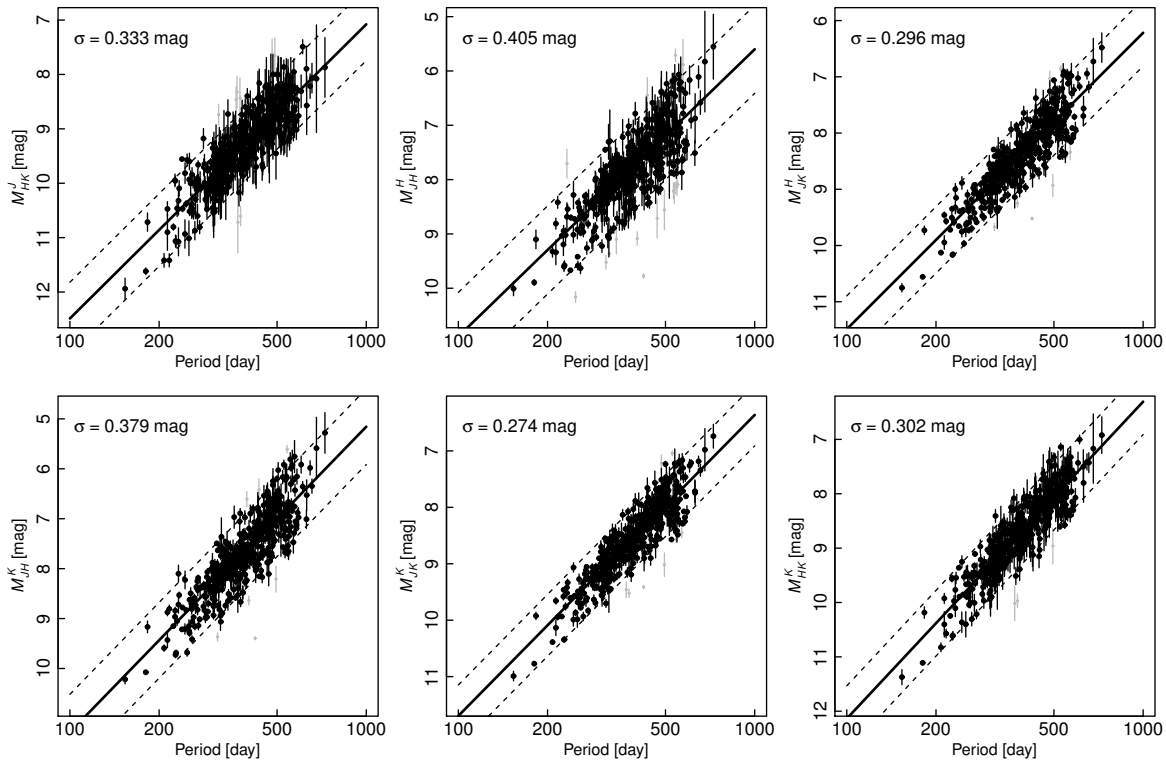

Figure 5.9: The oPLCRs of C-rich Miras in various combinations of bands and color terms in the LMC. The black solid lines indicate the best-fit of correlations of residuals, while the black dashed lines indicate the $\pm 2\sigma$ widths of the relations. Extreme outliers are indicated by gray points.

Table 5.4. NIR Observed PLCRs for C-Rich Miras

Index	b	a_0	a_1	σ
M^J_{HK}	3.78	10.86±0.04	-5.41±0.14	0.33
M^H_{JH}	2.89	9.30±0.05	-5.29±0.17	0.41
M^H_{JK}	1.41	9.91±0.04	-5.27±0.12	0.30
M^K_{JH}	2.20	9.44±0.05	-6.12±0.16	0.38
M^K_{JK}	0.91	10.10±0.04	-5.33±0.11	0.27
M^K_{HK}	1.98	10.39±0.04	-5.83±0.13	0.30

6. PROSPECTS FOR IDENTIFICATION OF OXYGEN-RICH MIRAS IN NEARBY GALAXIES WITH LSST

Miras are low- to intermediate-mass stars in their final stages of evolution and present in all types of galaxies. O-rich Miras are promising distance indicators as they exhibit relative tight PLRs at the NIR wavelengths. The upcoming Large Synoptic Survey Telescope (LSST) provides a great opportunity for discovering Miras in extragalactic systems. We estimated the discovery rates of O-rich Miras with LSST, using the simulated dataset from Yuan et al. (2017) and the expected cadence of observations by the LSST main survey. We found that our method will discover a considerable number of O-rich Miras in ~ 75 systems within 15 Mpc.

6.1 Miras in Nearby Galaxies with LSST

LSST is planning to carry out a wide-field, ground-based, optical survey and collect time-series observations from 2022 through 2032 (LSST Science Collaboration et al., 2009). With a primary mirror of 6.7m effective collecting area and a 3.5°-wide field of view, LSST will cover nearly half of the sky in the u, g, r, i, z and y bands with unprecedented cadence and depth. For this analysis, we selected a sample of 203 galaxies from the Extragalactic Distance Database (EDD; Tully et al., 2009) with $D < 15$ Mpc located within the LSST main survey footprint. The locations of these galaxies is shown in Figure 6.1.

We extracted the simulated riz observation times of the galaxies in the sample from the LSST OpSim database (Coffey et al., 2006) which takes into acount weather statistics for CTIO and the optimization strategy of LSST. The median number of measurements for these galaxies is 199, 202, 180 in r, i, and z, respectively. Since most measurements will be taken in pairs, the actual number of epochs is roughly half of the number of measurements.

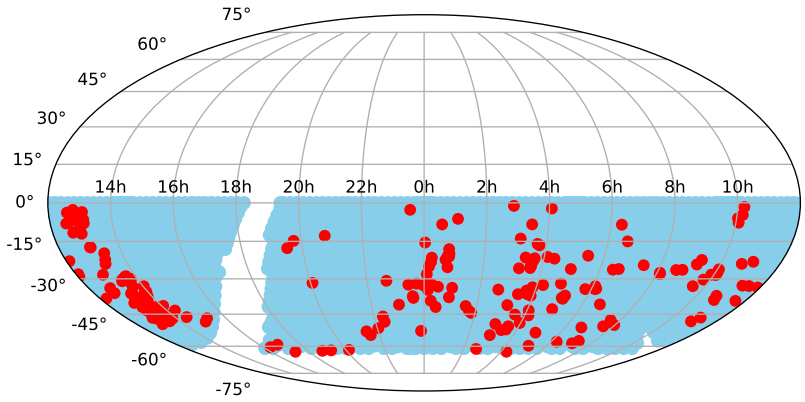

Figure 6.1: Sky coverage of the LSST main survey (blue shaded region) and the positions of the selected galaxies in the sample (red points). Within the distance of 15 Mpc, 203 galaxies found from EDD that are covered by the footprint of the LSST main survey.

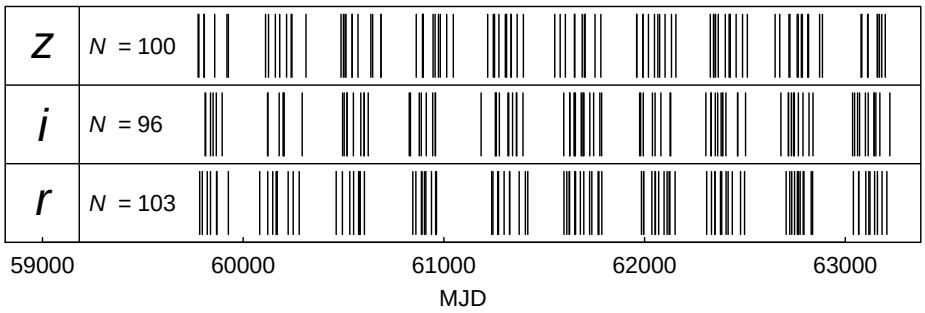

Figure 6.2: Simulated observation times for NGC 300 in riz and the total number of epochs. Any measurements taken within 0.1 day are considered as a single epoch.

Figure 6.2 shows the sampling pattern in riz for NGC 300 as an example. We did not include the observations in other bands in this study, as Miras will be too faint to be detected at those wavelengths.

We estimated the magnitude of O-rich Miras in riz as a function of period based on the brightest maximum I-band magnitude ($I_{\rm min}$) PLR and the period dependence of $V-I$ color for the LMC Mira sample. We used the $I_{\rm min}$ PLR obtained by Yuan et al. (2017)

$$I_{\rm min} = -4.83 - 2.12(\log P - 2.3) - 5.01(\log P - 2.3)^2 \tag{6.1}$$

and the period-color relation derived from the LMC sample (as shown in Figure 6.3)

$$V - I = 2.63 + 2.61(\log P - 2.3) \tag{6.2}$$

to derive the PLRs in riz bands. We obtained color transformations using SYNPHOT* and the Kurucz (1993) stellar models of giants at various temperatures. The best-fit transformations are

$$\begin{aligned} r - I &= 1.03 + 0.76(V - I - 1.5) + 0.19(V - I - 1.5)^2 \\ i - I &= 0.53 + 0.16(V - I - 1.5) + 0.04(V - I - 1.5)^2 \\ z - I &= 0.21 - 0.24(V - I - 1.5) - 0.02(V - I - 1.5)^2. \end{aligned} \tag{6.3}$$

Combining Equations 6.1, 6.2, and 6.3, we obtained

$$\begin{aligned} r_{\rm min} &= -2.61 + 0.96(\log P - 2.3) - 3.73(\log P - 2.3)^2 \\ i_{\rm min} &= -3.97 - 1.49(\log P - 2.3) - 4.77(\log P - 2.3)^2 \\ z_{\rm min} &= -4.77 - 2.65(\log P - 2.3) - 4.91(\log P - 2.3)^2. \end{aligned} \tag{6.4}$$

*http://www.stsci.edu/institute/software_hardware/stsdas/synphot

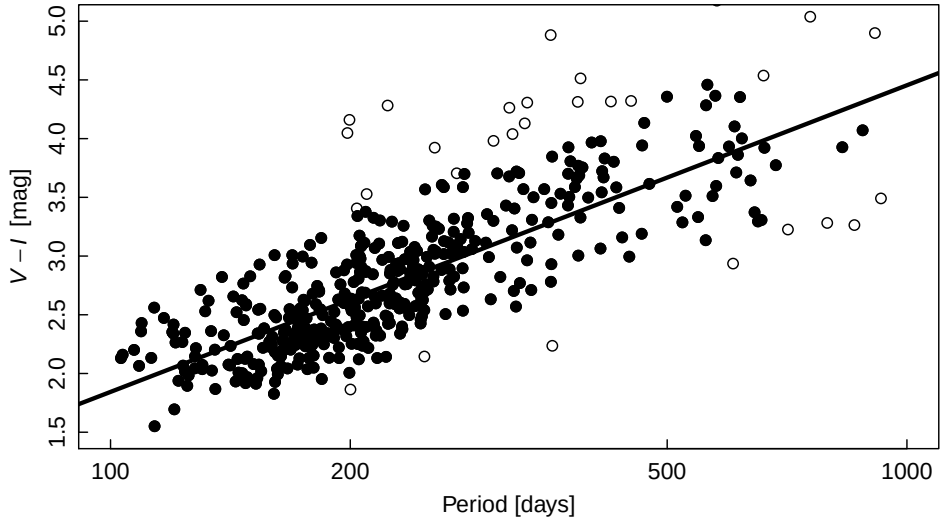

Figure 6.3: The color-period relation for the O-rich Miras in the LMC. $V - I$ is based on mean magnitudes obtained from Soszyński et al. (2009). The solid line indicates the best-fit and open circles are outliers.

Figure 6.4 shows these PLRs.

We obtained the expected 5 σ limiting magnitudes for LSST observations of these galaxies from the OpSim database. The median 5 σ limiting magnitudes are 24.2 mag, 23.4 mag, and 22.2 mag in r, i, and z, respectively. This is equivalent to a 500d-period O-rich Mira (at the maximum light) at a distance of roughly 2.5, 5.5, and 6 Mpc, respectively. However, the identification of Miras requires time-series measurements at different phases. We estimated the O-rich Mira discovery rate in these galaxies via the following calculations.

6.2 Relation Between Mira Discovery Rate and Light Curve Quality

We have recently developed a semi-parametric Gaussian process model to describe the Mira light curves and to search for their pulsation periods using sparsely-sampled light curves (He et al., 2016). We applied this model to M33 observations and used machine learning methods to discover over 1800 Miras in that galaxy (Yuan et al., 2017). We re-

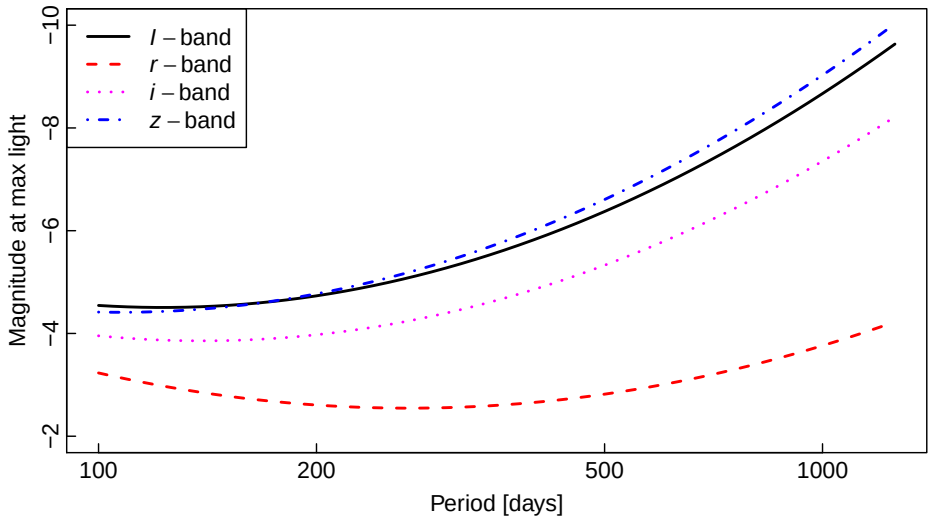

Figure 6.4: O-rich Mira PLRs at maximum light in r (red dashed curve), i (magenta dotted curve), and z (blue dash-dotted curve), based on tranformations from I (black solid curve) and the period-color dependence of Miras.

analyzed the simulated Mira light curves and results from this work to obtain an empirical relation between the signal-to-noise ratio (*SNR*) of a light curve around maximum light and the corresponding Mira detection rate.

We firstly computed the *SNR* for 10^4 simulated O-rich Mira light curves. For each light curve, we adopted the median uncertainty of the top 10% brightest "measurements" as the error σ_m of $I_{\rm min}$. The *SNR* was then calculated by $\frac{1.0875}{\sigma_m}$. We binned the simulated light curves based on *SNR* and computed the Mira discovery rates $Prob = \frac{N(P>0.5)}{N}$ in each bin, where $N(P > 0.5)$ is the number of simulated Miras with RF-based probability greater than 0.5, and N is the total number in that bin. Figure 6.5 shows the Mira discovery rates as a function of *SNR* at $I_{\rm min}$. We used the empirical function $Prob = 1 - \frac{3}{SNR}$ to describe the correlation, as shown by the solid curve in Figure 6.5. The relation between light curve quality and Mira detection rate is not associated with the distances of the host galaxies, and we can safely use it as an estimator of the Mira discovery rate.

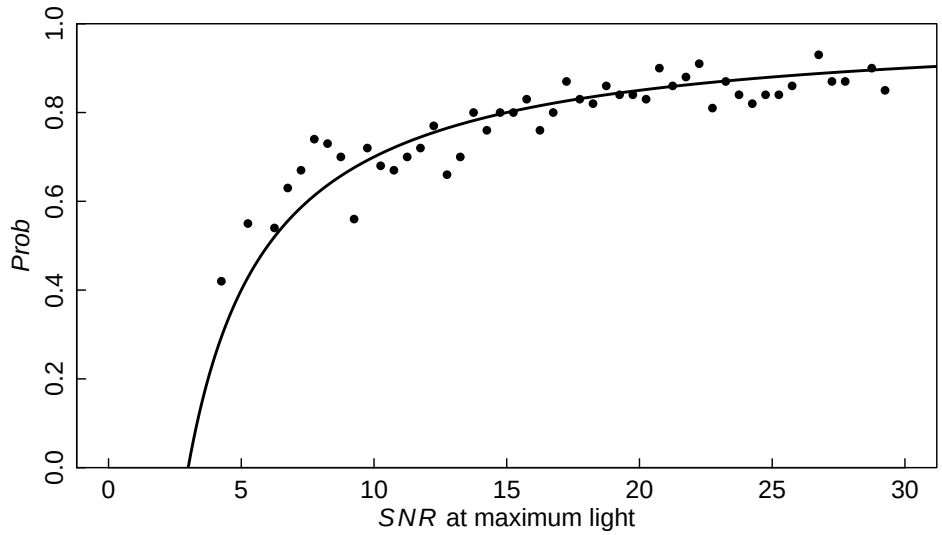

Figure 6.5: O-rich Mira detection rate against *SNR* at maximum light (black points) and the empirical relation between them (black curve).

6.3 Simulation of O-rich Mira Light Curves

We calculated the O-rich Mira discovery rate in the 203 galaxies of our sample by simulating light curves and computing their *SNR*s at maximum light. In this section, we give details of the simulation of light curves at various distances.

We derived the fiducial period distribution of O-rich Miras using the samples in the LMC and M33. The two galaxies have similar stellar masses (2×10^9 and $3 - 6 \times 10^9 M_\odot$, respectively; Kim et al., 1998; Corbelli, 2003). We averaged the O-rich Mira period distributions in the two galaxies with a 1:2 weighting to account for the differences in stellar masses, and adopted the number of O-rich Miras in M33 (1581) as the fiducial absolute amount. The fiducial distribution of O-rich Mira periods is shown in Figure 6.6.

O-rich Miras have a large range of amplitudes. In the simulation, we adopted the I-band amplitude distribution of the LMC O-rich Miras as the fiducial distribution for the amplitudes in riz bands. While this is a simplifying assumption, the result should only

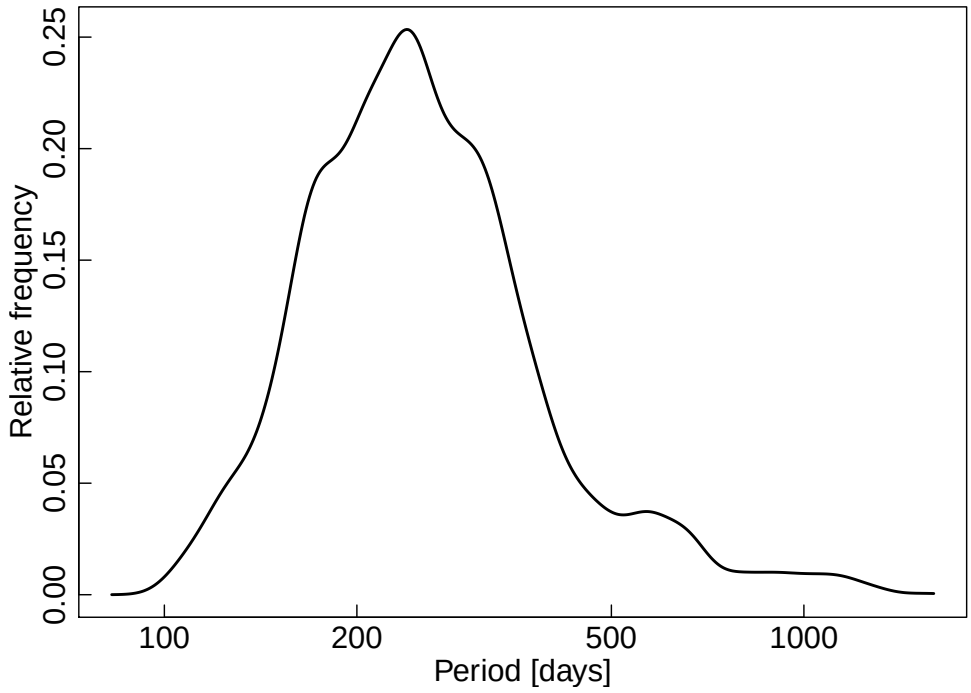

Figure 6.6: Fiducial distribution of O-rich Mira periods based on the LMC and M33 samples.

marginally depend on this distribution, given the large dispersion of Mira amplitudes as shown in Figure 6.7. We modeled the amplitude as a function of period with a Gaussian-like dispersion. The best-fit relation and 1 σ dispersion are shown in Figure 6.7 using solid and dashed lines, respectively. In the simulation, we randomly drew amplitudes from this relation but rejected any amplitudes less than 1 mag or greater than 5 mag.

For each galaxy, we simulated 240 light curves with periods P drawn from the fiducial period distribution. We adopted sinusoidal curves

$$m = m_{\min} + 0.5A + 0.5A \cdot \sin(2\pi\phi) + \mu \tag{6.5}$$

where A is drawn from the fiducial amplitude distribution, μ is the distance modulus of the galaxy, $\phi = \frac{(t-t_0) \mod P}{P}$ is the phase and the reference time t_0 is randomly drawn from a

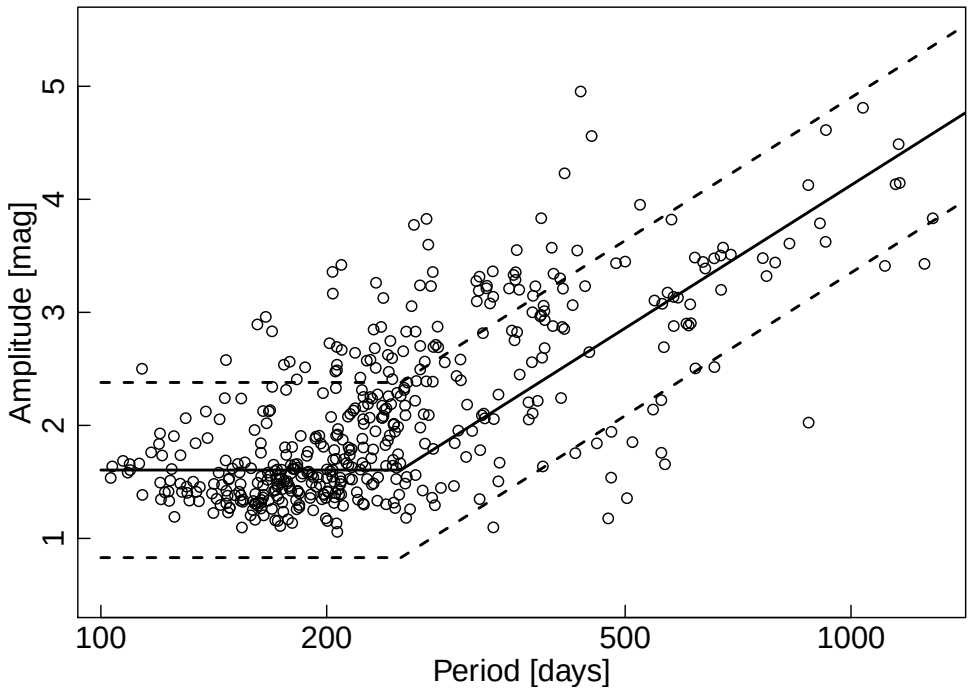

Figure 6.7: Fiducial distribution of O-rich Mira amplitudes based on the LMC I-band light curves.

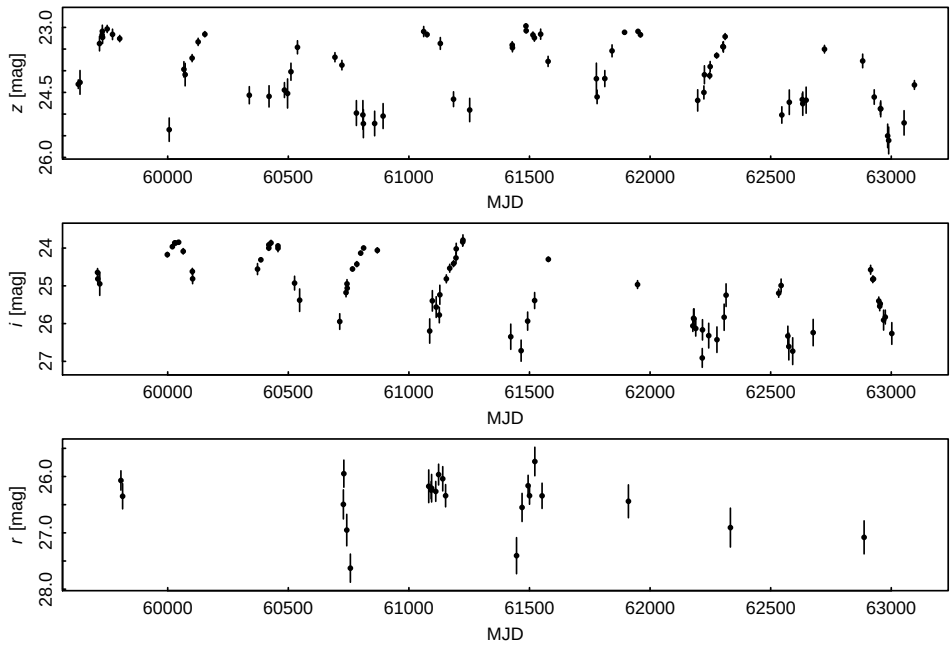

Figure 6.8: Simulated riz light curves for a 430d-period Mira in NGC 5398, based on the expected cadence of LSST.

uniform distribution. The observation time t is the simulated observation time for a given galaxy extracted from OpSim. For measurements with a time span less than 0.1 day, we combined them as a single epoch and adjusted the *SNR* accordingly. Figure 6.8 shows an example of simulated Mira light curve with a 430d period in NGC 5398.

6.4 Results

6.4.1 SNR of the Simulated LSST Measurements

We calculated the *SNR* of individual simulated LSST measurement at given time using the following simple model:

$$SNR = \frac{S}{\sqrt{S + B + I}} \tag{6.6}$$

$$S = 10^{-0.4(m - m_{zp} - k(1-X))} \times g \times \Delta t \tag{6.7}$$

$$B = 10^{-0.4(m_{sky} - m_{zp})} \times g \times \Delta t \times A \tag{6.8}$$

$$I = 12.7 \times A \tag{6.9}$$

$$A = p^2 \times F^2 \times 2.266 \tag{6.10}$$

where S is the counts from the source, B is the counts from the sky background, I is the counts from instrument noise, m is the source magnitude, m_{zp} is the zeropoint, k is the atmospheric extinction coefficient, X is the airmass, g is the gain of the instrument, Δt is the exposure time, A is the area of aperture, p is the pixel scale, and F is the FWHM. We list the values for the time-independent variables in Table 6.1. Values for the time-dependent variables were extracted from the OpSim database.

We tested the model with the 5 σ limiting magnitudes extracted from OpSim, which in principle have *SNR*s equal to 5. We found our simple model gives *SNR*s of 5.5, 5.3, and 5.15 in r, i, and z, respectively. Thus, we adjusted the coefficient in Equation 6.10 to be 2.763 for r, 2.545 for i, and 2.407 for z.

Table 6.1. Values of Time-Independent Variables in *SNR* Model

Variable	Meaning	value	Units
r_{zp}	r-band zeropoint	27.23	mag
i_{zp}	i-band zeropoint	26.89	mag
z_{zp}	z-band zeropoint	26.50	mag
k_r	r-band atmospheric extinction coef.	0.13	mag
k_i	i-band atmospheric extinction coef.	0.10	mag
k_z	z-band atmospheric extinction coef.	0.07	mag
g	gain	2.3	e^-/ADU
p	pixel scale	0.2	$''$/pix

6.4.2 Discovery Rates of O-rich Miras with LSST

For each simulated light curve, we calculated the probability of being detected with $P_i = 1 - \frac{3}{SNR}$. To be conservative, we added another constrain on the detection probability that there must be at least 20 epochs with $SNR > 3$. The discovery rate is obtained by summing probalities over all the simulated light curves $R = \frac{\sum_{i=1}^{240} P_i Q_i}{240}$, where Q_i is equal to 1 if there are at least 20 epochs with $SNR > 3$, or is equal to 0 otherwise.

We estimated the total number of O-rich Miras that can be discovered in each galaxy by scaling the discovery rates with the ratio of the K_s luminosity of the host galaxy to that of M33. As an example, Figure 6.9 shows the number of O-rich Miras in NGC 4802 can be discovered by LSST and the number of epochs with $SNR > 3$ over ten years as a function of Mira period. Based on our estimations, there are 75 galaxies in which over 100 O-rich Miras can be discovered with LSST, with the farthest one at a distance of 14 Mpc. Figure 6.10 shows the histogram of the expected number of Miras that will be discovered with LSST across all the galaxies. With the observaing strategy presented in OpSim, we

will be able to discover O-rich Miras two years after the survey start, and discover 2×10^5 O-rich Miras in the end, as shown in Figure 6.11.

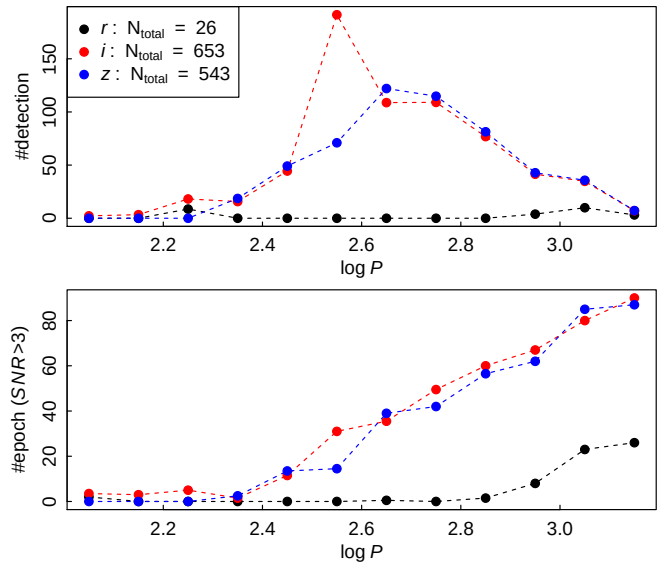

Figure 6.9: Number of expected O-rich Miras to be detected by LSST (upper) and number of epochs with $SNR > 3$ (lower) as a function of Mira period in r (black), i (red), and z (blue) bands for the galaxy NGC 4802.

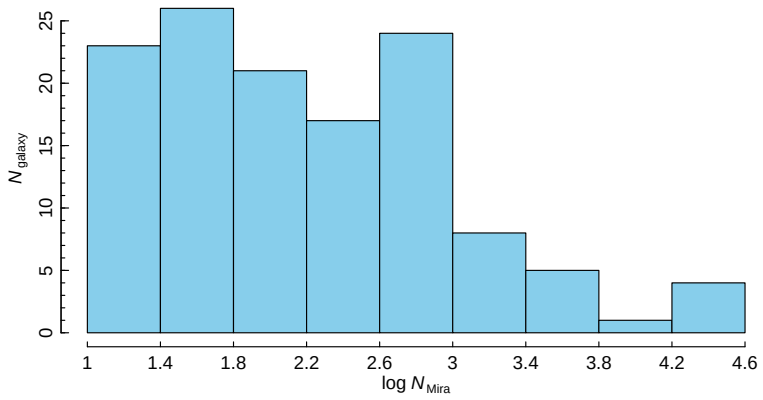

Figure 6.10: Histogram of expected number of O-rich Miras that will be discovered with LSST across all the galaxies in the sample.

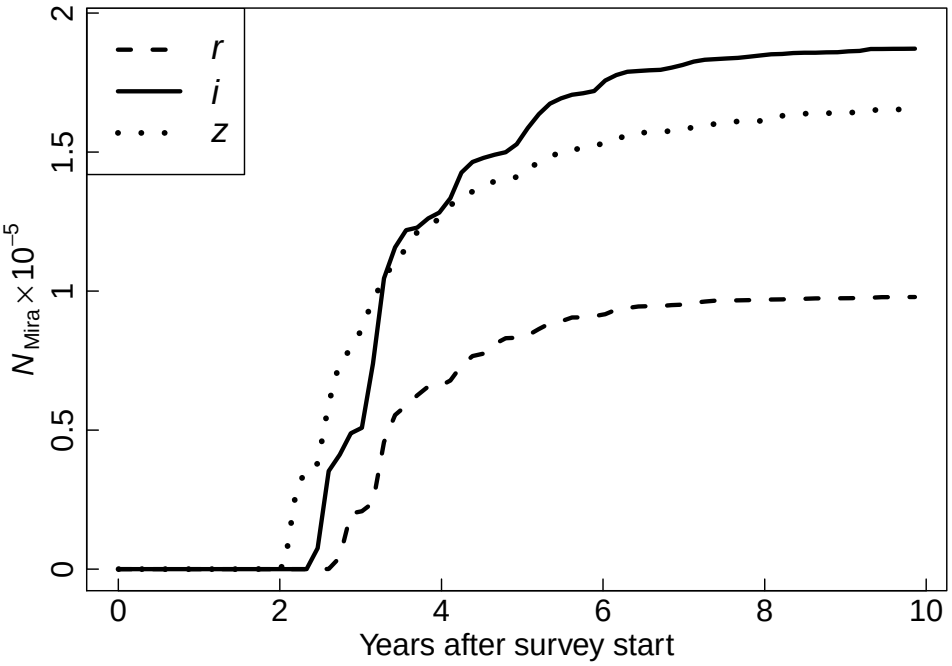

Figure 6.11: Total number of Miras expected to be discovered by LSST in riz as a function of time since the start of the survey.

7. SUMMARY

We have presented techniques for detecting Cepheids and Miras in extragalactic systems. We discovered over 800 Cepheid candidates in M101, a subsample of which contributed to a recent measurement of H_0. We identified over 1800 Mira candidates in M33 with the aid of machine learning techniques. We determined the phases of 34 Galactic Cepheids, which will be used to correct *HST* observations to mean light. We studied the NIR properties of the LMC Mira candidates and obtained PLRs for the O-rich subtype.

We highlight the usefulness of machine learning techniques in searching for variable sources among large number of unclassified objects. In the era of big data, it is not feasible to carry out visual inspection of individual objects, and machine learning can take its place. With random forest, a tree-based machine learning method for classification, we not only identified Mira candidates among $\sim 240,000$ objects, but also were able to estimate the recovery rate with cross-validations. We have demonstrated that our method is capable of identifying Miras using the data products from LSST.

We also highlight O-rich Miras as distance indicators. By deriving the mean magnitudes of Miras with time-series observations of the LMC Mira candidates, we have confirmed that the O-rich Mira PLR at K_s-band is almost as tight as Cepheid PLR at the same band. Moreover, Miras as distance indicators have two advantages over Cepheids: (1) They are ~ 2 mag brighter than Cepheids in NIR wavelengths. (2) Miras are old stars and present in all types of galaxies. With these properties, Miras are promising distance indicators.

REFERENCES

Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, MNRAS, 459, 2408

Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 43, Sky Surveys. Protostars to Protogalaxies, ed. B. T. Soifer, 291

Alibert, Y., Baraffe, I., Hauschildt, P., & Allard, F. 1999, A&A, 344, 551

Aringer, B., Girardi, L., Nowotny, W., Marigo, P., & Lederer, M. T. 2009, A&A, 503, 913

Baranowski, R., Smolec, R., Dimitrov, W., et al. 2009, MNRAS, 396, 2194

Barsukova, E. A., Goranskij, V. P., Hornoch, K., et al. 2011, MNRAS, 413, 1797

Benedict, G. F., McArthur, B. E., Feast, M. W., et al. 2007, AJ, 133, 1810

Bonanos, A. Z., Stanek, K. Z., Kudritzki, R. P., et al. 2006, ApJ, 652, 313

Bresolin, F. 2007, ApJ, 656, 186

—. 2011, ApJ, 730, 129

Cami, J., Bernard-Salas, J., Peeters, E., & Malek, S. E. 2010, Science, 329, 1180

Carpenter, J. M. 2001, AJ, 121, 2851

Casertano, S., Riess, A. G., Anderson, J., et al. 2016, ApJ, 825, 11

Coffey, C. R., Saha, A., & Miller, M. 2006, in Bulletin of the American Astronomical Society, Vol. 38, American Astronomical Society Meeting Abstracts, 1021

Collins, G. W. 1989, The fundamentals of stellar astrophysics

Corbelli, E. 2003, MNRAS, 342, 199

DePoy, D. L., Atwood, B., Belville, S. R., et al. 2003, "SPIE Conf. Ser., 4841, 827

Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, MNRAS, 414, 2602

Einstein, A. 1916, Annalen der Physik, 354, 769

Faraway, J., Mahabal, A., Sun, J., et al. 2016, Statistical Analysis and Data Mining: The

ASA Data Science Journal, Vol. 9, Issue 1, p. 1-11, 9, 1

Feast, M. W. 2009, in AGB Stars and Related Phenomena, ed. T. Ueta, N. Matsunaga, & Y. Ita, 48

Fernie, J. D. 1984, in IAU Symposium, Vol. 105, Observational Tests of the Stellar Evolution Theory, ed. A. Maeder & A. Renzini, 441

Fouqué, P., Arriagada, P., Storm, J., et al. 2007, A&A, 476, 73

Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, ApJ, 553, 47

Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377

Gerasimovic, B. P. 1928, Proceedings of the National Academy of Science, 14, 963

Gieren, W. P., Moffett, T. J., & Barnes, III, T. G. 1999, ApJ, 512, 553

Glass, I. S., & Lloyd Evans, T. 1981, Nature, 291, 303

—. 2003, MNRAS, 343, 67

Gonzaga, S., & et al. 2012, The DrizzlePac Handbook

Guzman-Ramirez, L., Lagadec, E., Wesson, R., et al. 2015, MNRAS, 451, L1

He, S., Yuan, W., Huang, J. Z., Long, J., & Macri, L. M. 2016, AJ, 152, 164

Hoffmann, S. L., Macri, L. M., Riess, A. G., et al. 2016, ApJ, 830, 10

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Humphreys, E. M. L., Reid, M. J., Moran, J. M., Greenhill, L. J., & Argon, A. L. 2013, ApJ, 775, 13

Inno, L., Matsunaga, N., Romaniello, M., et al. 2015, A&A, 576, A30

Ita, Y., & Matsunaga, N. 2011, MNRAS, 412, 2345

Javadi, A., Saberi, M., van Loon, J. T., et al. 2015, MNRAS, 447, 3973

Kanbur, S. M., Hendry, M. A., & Clarke, D. 1997, MNRAS, 289, 428

Kholopov, P. N., Samus, N. N., Kazarovets, E. V., & Perova, N. B. 1985, Information Bulletin on Variable Stars, 2681

Kim, S., Staveley-Smith, L., Dopita, M. A., et al. 1998, ApJ, 503, 674

Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in Proc. SPIE, Vol. 8127, Optical Modeling and Performance Predictions V, 81270J

Kurucz, R. L. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 44, IAU Colloq. 138: Peculiar versus Normal Phenomena in A-type and Related Stars, ed. M. M. Dworetsky, F. Castelli, & R. Faraggiana, 87

Laney, C. D., & Stobie, R. S. 1993, MNRAS, 260, 408

Leavitt, H. S., & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1

Lemaître, A. G. 1931, Nature, 128, 704

Lomb, N. R. 1976, Ap&SS, 39, 447

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201

Ludendorff, H. 1928, Handbuch der Astrophysik, 6, 49

Macri, L. M., Ngeow, C.-C., Kanbur, S. M., Mahzooni, S., & Smitka, M. T. 2015, AJ, 149, 117

Macri, L. M., Stanek, K. Z., Sasselov, D. D., Krockenberger, M., & Kaluzny, J. 2001, AJ, 121, 861

Madore, B. F. 1982, ApJ, 253, 575

Madore, B. F., & Freedman, W. L. 1991, PASP, 103, 933

Massey, P., Olsen, K. A. G., Hodge, P. W., et al. 2006, AJ, 131, 2478

Mattei, J. A. 1997, Journal of the American Association of Variable Star Observers (JAAVSO), 25, 57

Meixner, M., Gordon, K. D., Indebetouw, R., et al. 2006, AJ, 132, 2268

Mink, D. J. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 172, Astronomical Data Analysis Software and Systems VIII, ed. D. M. Mehringer, R. L. Plante, & D. A. Roberts, 498

Miyata, T., Kataza, H., Okamoto, Y., Onaka, T., & Yamashita, T. 2000, ApJ, 531, 917

Monson, A. J., & Pierce, M. J. 2011, ApJS, 193, 12

Mutschke, H., Andersen, A. C., Clément, D., Henning, T., & Peiter, G. 1999, A&A, 345, 187

Ngeow, C.-C., Kanbur, S. M., Neilson, H. R., Nanthakumar, A., & Buonaccorsi, J. 2009, ApJ, 693, 691

Nishiyama, S., Tamura, M., Hatano, H., et al. 2009, ApJ, 696, 1407

Pellerin, A., & Macri, L. M. 2011, ApJS, 193, 26

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Persson, S. E., Madore, B. F., Krzemiński, W., et al. 2004, AJ, 128, 2239

Pietrzyński, G., Graczyk, D., Gieren, W., et al. 2013, Nature, 495, 76

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Pojmanski, G. 1997, AcA, 47, 467

Rajan, A., & et al. 2011, WFC3 Data Handbook v. 2.1

Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, ApJ, 733, 10

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Riess, A. G., Macri, L., Casertano, S., et al. 2009, ApJ, 699, 539

—. 2011, ApJ, 730, 119

Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56

Romaniello, M., Primas, F., Mottini, M., et al. 2008, A&A, 488, 731

Samus, N. N., Kazarovets, E. V., Durlevich, O. V., Kireeva, N. N., & Pastukhova, E. N. 2017, Astronomy Reports, 61, 80

Scargle, J. D. 1982, ApJ, 263, 835

Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, A&AS, 96, 269

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Sirianni, M., Jee, M. J., Benítez, N., et al. 2005, PASP, 117, 1049

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Smolec, R., & Moskalik, P. 2008, AcA, 58, 193

Soszyński, I., Poleski, R., Udalski, A., et al. 2008, AcA, 58, 163

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009, AcA, 59, 239

—. 2011, AcA, 61, 217

—. 2013, AcA, 63, 21

Stetson, P. B. 1987, PASP, 99, 191

Stetson, P. B. 1993, in IAU Colloq. 136: Stellar Photometry - Current Techniques and Future Developments, ed. C. J. Butler & I. Elliott, Vol. 136, 291

—. 1994, PASP, 106, 250

—. 1996, PASP, 108, 851

Templeton, M. R., & Karovska, M. 2009, ApJ, 691, 1470

Thompson, T. A., Prieto, J. L., Stanek, K. Z., et al. 2009, ApJ, 705, 1364

Tibshirani, R. 1994, Journal of the Royal Statistical Society, Series B, 58, 267

Tully, R. B., Rizzi, L., Shaya, E. J., et al. 2009, AJ, 138, 323

Turner, D. G. 1996, JRASC, 90, 82

Udalski, A., Szymanski, M. K., Soszynski, I., & Poleski, R. 2008, AcA, 58, 69

van Belle, G. T., Thompson, R. R., & Creech-Eakman, M. J. 2002, AJ, 124, 1706

van Leeuwen, F. 2007, A&A, 474, 653

van Leeuwen, F., Feast, M. W., Whitelock, P. A., & Laney, C. D. 2007, MNRAS, 379, 723

van Maanen, A. 1928, Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington, 356, 1

Wang, S., & Jiang, B. W. 2014, ApJL, 788, L12

Whitelock, P. A., Feast, M. W., Marang, F., & Groenewegen, M. A. T. 2006, MNRAS, 369, 751

Whitelock, P. A., Feast, M. W., Marang, F., & Overbeek, M. D. 1997, MNRAS, 288, 512

Whitelock, P. A., Feast, M. W., & van Leeuwen, F. 2008, MNRAS, 386, 313

Wood, P. R. 1990, in Astronomical Society of the Pacific Conference Series, Vol. 11, Confrontation Between Stellar Pulsation and Evolution, ed. C. Cacciari & G. Clementini, 355–363

Wood, P. R., Alcock, C., Allsman, R. A., et al. 1999, in IAU Symposium, Vol. 191, Asymptotic Giant Branch Stars, ed. T. Le Bertre, A. Lebre, & C. Waelkens, 151

Yoachim, P., McCommas, L. P., Dalcanton, J. J., & Williams, B. F. 2009, AJ, 137, 4697

Yuan, W., He, S., Macri, L. M., Long, J., & Huang, J. Z. 2017, AJ, 153, 170